Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
HERSCHEL W. LEIBOWITZ, *International Business Machines Corporation, Bethesda, Maryland*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Set, Suggestion, and Conditioning: W. W. GRINGS, S. CARLIN, AND M. H. APPLEY 417
Acuity and the Statistical Theory of Figural Aftereffects: F. H. GEORGE 423
Assimilation vs. Contrast in the Anchoring of Perceptual Judgments of Weight: A. PARDUCCI AND L. M. MARSHALL 426
Cue Neutralization: The Effects of Random Reinforcements upon Discrimination Learning: M. LEVINE 438
Reversal and Nonreversal Shifts in Discrimination Learning in Retardates: B. HOUSE AND D. ZEAMAN 444
The Cue Additivity Principle in a Restricted Social Interaction Situation: J. M. RICHARDS, JR. 452
Response-Mediated Generalization with Simple Skeletal-Motor Responses: S. H. SIMON 458
Escape Performance as a Function of Delay of Reinforcement: H. FOWLER AND M. A. TRAPOLD 464
Response Strength as a Function of Drive Level and Pre- and Postshift Incentive Magnitude: D. EHRENFREUND AND P. BADIA 468
The Induction of Nonveridical Slant and the Perception of Shape: W. EPSTEIN, H. BONTRAGER, AND J. PARK 472
Retention of Connected Discourse as a Function of Duration of Interpolated Learning: N. J. SLAMECKA 480
Effects of Instructional Set and UCS Intensity on the Latency, Percentage, and Form of the Eyelid Response: I. GORMEZANO AND J. W. MOORE 487
Tendencies to Eye Movement and Perceptual Accuracy: H. F. CROVITZ AND W. DAVES 495
The Effect of Nonsense-Syllable Compound Stimuli on Latency in a Verbal Paired-Associate Task: B. S. MUSGRAVE 499
Combination of Drive and Drug Effects: J. MENDELSON AND D. BINDRA 505
Anagram Solution Times: A Function of Word Transition Probabilities: M. S. MAYZNER AND M. E. TRESSELT 510
Transition Probability Effects in Anagram Problem Solving: H. BEILIN AND R. HORN 514
Supplementary Report: A Nonverbal Measure of Extinction in Skill and Chance Situations: K. B. HOLDEN AND J. B. ROTTER 519

American Psychological Association

Vol. 63, No. 5 May 1962

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $10.00, or $20.00 annually. Single copies are $2.00. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.* When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1961 the publication cost is $30 per printed page.

All **regular** articles, **supplementary reports,** and **replication reports** should be submitted to the Editor-elect: David A. Grant: Department of Psychology, University of Wisconsin; Madison 6, Wisconsin. **Early publication** articles should be sent to the Editor: Arthur W. Melton; Department of Psychology, University of Michigan; Ann Arbor, Michigan until *July 1, 1962.* Thereafter they should be sent to the Editor-elect. Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts.* Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR
Acting Production Manager

RUTH P. JAGOLINZER
Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1962 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 63, NO. 5 MAY 1962

SET, SUGGESTION, AND CONDITIONING[1]

WILLIAM W. GRINGS, SIDNEY CARLIN
University of Southern California
AND MORTIMER H. APPLEY
Southern Illinois University

The concept of reinforcement, as it applies to classical conditioning of autonomic responses with human *S*s, has several different interpretations. To many workers the reference of the term is strictly empirical, whereas to other workers there is strong implication that the essential requirement is a change in a motivational state (Mowrer, 1960; Spence, 1956, 1960).

Some theorists have emphasized reinforcement as a *reactive* process within *S*. Woodworth (1947, 1958), for example, asserts that during conditioning *S* is learning to anticipate stimulation. Conditioning involves essentially "a change in his (*S*'s) way of receiving, or perceiving, the sequence of stimuli and especially the preliminary or conditioned stimulus" (Woodworth, 1947, p. 122).

Variations on the notions of reinforcement as some form of confirmation of an anticipatory reaction by *S* to the CS are numerous. Quite different discussions are provided by Hilgard and Marquis (1940), Bruner (1951), Anokhin (1955), and Solley and Murphy (1960). One common weakness of most "reinforcement of perception" concepts is that they do not have adequate operational components. The integration of concepts of perception and reinforcement was the subject of a paper by Razran (1955). And one of the present authors (Grings, 1960) has proposed operations for inferring a preparatory set concept in the context of classical conditioning.

The present experiment was set up to explore some of the implications of a reinforcement concept based on an assumed anticipation of stimulation. It was an attempt to demonstrate, with human *S*s, that a tendency to anticipate stimulation (i.e., a response set) can be manipulated in a situation analogous to classical conditioning. The operations for establishing the set were suggested by Cornbecker, Welch, and Fisichelli (1949) in a study of conditioning factors underlying hypnosis.

The conditioned response (CR) was defined as the GSR to a two-word

[1] This research was supported by a grant (G-8612) from the National Science Foundation.

verbal cue (CS) capable of suggesting stimulation. Two test cues, ELECTRIC SHOCK and SCALDING WATER, were considered initially (intrinsically) adequate to elicit GSRs, and a control cue PLEASANT DAY was used for reference purposes. The two test cues differed, however, with respect to their adequacy to suggest stimulation in the particular experimental setting. This difference was labeled cue *feasibility*. ELECTRIC SHOCK was called feasible since an *S* with GSR and ground electrodes attached might reasonably expect an actual shock. On the other hand, the cue SCALDING WATER was called unfeasible since the otherwise barren experimental cubicle contained no apparent means for presenting the actual stimulus. (The neutral cue PLEASANT DAY was assumed not to be suggestive of actual stimulation.)

Changes in GSR to the three test cues were studied as a function of two independent variables: feasibility, as defined above; and amount of reinforcement. Reinforcement was not employed in the usual sense of the term, for the reinforcing state of affairs (presentation of a UCS) was not applied directly to the test cues previously mentioned. That is, the cue ELECTRIC SHOCK was never actually followed by a shock, or SCALDING WATER followed by scalding water. Instead, seven other cues were presented and were followed by appropriate stimulation.

The single sequence of one verbal cue followed by the stimulation it suggested was arbitrarily labeled as one reinforcement. As experimental variation two groups were used: a *long* training group received seven series of the seven cues plus a test series; the other, a *short* training group, received two series of the training cues plus a test series.

METHOD

Subjects.—Eighty paid volunteers, 49 men and 31 women, from undergraduate psychology classes at the University of Southern California were observed for periods of from 45 min. to 90 min.

Apparatus.—The GSR was obtained as a dc resistance through a zinc, saline paste, skin junction employing ¾-in. diameter Lucite cup electrodes fastened by rubber straps to the palm and back of *S*'s nonpreferred hand. The electrodes led to a modified Darrow-type bridge the output of which was amplified by a Brush, Model BL 530 dc amplifier and recorded on an Esterline-Angus ink writing milliameter. Current through *S* was 45 μa.

The *S*s were seated at a table in one section of the laboratory. By appropriate switching in another portion of the laboratory *E* could cause (*a*) an 8-in. fan to blow across *S*, (*b*) a green flood light to come on, (*c*) tape recorded symphonic music to play, (*d*) all room lights to go out, (*e*) a 100-db., 1000°-cps tone to sound, (*f*) a light bulb in front of *S* to blink at the rate of 10/sec, or (*g*) a mechanical transducer on *S*'s finger to vibrate. The *E* could also project verbal cues from a 2 × 2 in. projector onto an 8 × 10 in. ground glass screen in front of *S*. The verbal cues corresponding to the forementioned stimulations were respectively (*a*) COOL BREEZE, (*b*) GREEN LIGHT, (*c*) SOFT MUSIC, (*d*) DARK ROOM, (*e*) LOUD TONE, (*f*) FLICKERING LIGHT, and (*g*) FINGER VIBRATION.

Verbal cues were presented for 10 sec. per trial. During the second half of this period the appropriate UCS was presented. Timing of this sequence was controlled by electronic interval timers. Intertrial intervals were varied randomly among 25, 30, and 35 sec.

General design.—As originally conceived the study was to employ four different treatment groups of 10 *S*s each in a 2 × 2 factorial arrangement with feasible vs. unfeasible test stimuli on one axis and long vs. short training conditions on the other axis. The first 40 *S*s were tested in this arrangement. However, a second 40 *S*s were tested, duplicating this condition, with the exception that instead of stopping the reinforcement of the *short* group after it was tested following 14 reinforcements, this group was given 35 additional reinforcements and tested again at the *long* test point. This made it possible to use 30 different *S*s for evaluation of certain cell means.

General procedure.—The *Ss* were told "This is a study of your palmar skin resistance (the sweat gland activity of the palms of your hands)." Then after a brief discussion of the GSR they were instructed to "Watch the screen and read aloud the words that will appear on it from time to time."

The *Ss* were first assigned by a table of random numbers to one of the four experimental groups: (*a*) short training series, feasible test cue; (*b*) short series, unfeasible test cue; (*c*) long series, feasible test cue; and (*d*) long series, unfeasible test cue. Then all were presented a pretest series which included all subsequently used cues and 10 additional "filler" cue phrases.

The short reinforcement condition contained two series of the seven reinforced cues plus a test series which contained the sequence of reinforced cues with test cues introduced at Ordinal Positions 1 and 6 with the series ending in a disparity trial, i.e., the verbal cue GREEN LIGHT being followed by the actual loud tone. The disparity trial was omitted from the short group during testing of the last 40 *Ss* since these *Ss* received further training and test following 35 additional training trials. For any particular *S* the two test cues would be either the feasible or unfeasible cue and the reference cue. The order of presentation (between Positions 1 and 6) was varied systematically.

The long reinforcement condition contained seven series of the seven reinforced cues plus the test series, as described above. Here, too, the test series ended in a disparity trial.

Following the test series, *Ss* were given a questionnaire to fill out. It posed such questions as "Did you expect to feel scalding water?" "Did you experience scalding water?" "Did you expect an electric shock?" "Did you experience an electric shock?"

RESULTS AND DISCUSSION

Table 1 shows the means and standard deviations of the magnitudes of response to the three critical or test cues as a function of the position in the reinforcement series. The basic unit of response is the square root of the maximum change in conductance, in micromhos, occurring during the 10 sec. following presentation of the cue.

A few review comments may help the interpretation of the data in this table. All *Ss* were tested at each initial point, hence the *N*s for comparisons of pretest response magnitudes are 80. When pretest comparisons are made, the response to suggested shock is significantly greater than the response to either of the other stimuli. Because of this initial difference among stimuli, it is the *change* in magnitude of response to these cues that is emphasized.[2] When changes are evaluated the *N*s at each point equal 30. These samples are composed of 10 *Ss* from the first replication of the study and 20 *Ss* from the second replication at which time *Ss* were tested at both the long and short training points.[3]

TABLE 1

GSRs (MICROMHOS) TO TEST CUES AT DIFFERENT LEVELS OF TRAINING

Cues	Training Level		
	Pre-training	14 Reinforcements	49 Reinforcements
ELECTRIC SHOCK			
Mean	1.03	1.34	1.14
SD	.69	.72	.67
N	80	30	30
SCALDING WATER			
Mean	.79	.94	.64
SD	.59	.61	.53
N	80	30	30
PLEASANT DAY			
Mean	.73	.64	.59
SD	.55	.53	.53
N	80	60	60

[2] In subsequent experimentation with these same stimuli, the initial differences among stimuli have been removed by administration of an adaptation series at the outset of the experiment.

[3] A comparison of the responses at the late training point, for *Ss* who had the interposed test at the early training point and *Ss* who had none, showed no difference. Data from the last "replication" of the experiment are therefore used at both the early and late training test points. Measures of change

TABLE 2

MEAN ADJUSTED SCORES FOR TEST STIMULI DURING PRETRAINING AND POST-ACQUISITION TEST PERIODS

Cue	Short Reinforcement Series		Long Reinforcement Series	
	Pre-training	Post-test	Pre-training	Post-test
Feasible	.15	.75	.37	.58
Unfeasible	.01	.20	.02	.22

Note.—N = 80; 20 *Ss* in each subgroup.
Difference in magnitude of GSR to individual test and reference stimuli measured in micromhos, square root of conductance change.

When changes are evaluated as the mean difference from pretraining to posttraining response levels, the change for the feasible cue is the only one which is statistically significant ($t = 2.76$; $df = 29$). That is, the early training experience increases the response evocation capability of the feasible cue. There also appears to be an early training increase for the unfeasible cue, but it is not a statistically significant one. Somewhere after this first test point the response evocation falls, for when tested at the late training point neither stimulus is significantly different from the pretraining value. It will be noted that the neutral or reference stimulus shows a consistent course of adaptation throughout the experiment. It is therefore quite probable that the training series was so long that general response adaptation occurred, competing strongly with the learning trends presumably operating early in training.

To confirm that there was adaptation, responses to the reinforced cues (LOUD TONE, COOL BREEZE, etc.) were plotted for successive trials. A typical curve of decreasing magnitude of response with repeated trials was obtained.

One way of partialing out general and situational adaptation is to subtract from the GSR elicited by the crucial test stimulus the GSR given to the neutral or reference stimulus on the closest adjacent trial. This is based on the assumption that the response to the test stimulus represented effects of both learning and adaptation whereas the response to the reference stimulus reflected only general situational adaptation. Tabulations and analyses were made using this adjusted score (the difference between response to test and reference stimuli) as the dependent variable. Mean values for the major experimental categories are given in Table 2.

To evaluate the significance of the differences evident in this table an analysis of variance as a mixed factorial design was performed. Feasibility and reinforcement are between-subject main effects and pretest vs. posttest periods provide a within-subject main effect. Three components are significant beyond the 1% level, as follows. The overall feasibility variable is significant ($F = 7.89$; $df = 1/76$). Here, it is important to recall that this test reflects pretest as well as posttest differences between feasible and unfeasible stimuli.

The overall change from pretest to posttest conditions is also significant ($F = 12.77$; $df = 1/76$); whereas, the difference between the two posttest groups (long reinforcement vs. short reinforcement) is not significant. The triple interaction of feasibility, reinforcement groups, and pre-to-posttest periods is also significant ($F = 14.22$; $df = 1/76$).

In summary, the analysis of variance on difference scores confirms the analysis of response means of Table 1. The principal variation

due to training were made on mean differences of samples of 30 *Ss* for each point on the curve.

occurs for the feasible stimulus cue during the short series of reinforcements.

To further evaluate the extent of general situational adaptation a second form of response measure, the perceptual disparity response was used. This involves presenting a CS followed by a UCS which differs in some specified way from the UCS with which the CS had been paired previously. In this case the cue GREEN LIGHT was presented, then at the end of 5 sec. the actual loud tone was given. The disparity response was computed as the difference in response to the actual tone UCS when it was preceded by the wrong or disparate cue GREEN LIGHT and the last previous response to the tone UCS when it was preceded by the cue LOUD TONE. The fact that this perceptual disparity response (for the 40 *Ss* on whom it was available) was significantly larger for the long reinforcement groups than for the short reinforcement groups was taken as further evidence that a significant degree of adaptation had occurred between 14 and 49 reinforcements. The data on this comparison have been reported elsewhere (Grings, 1960).

It had been predicted at the start of the study that the tendency to respond to an unfeasible cue would show up more strongly with increased training. In other words, the longer the reinforcement series, the stronger the set might be to expect stimulation on the scalding water cue. This prediction was not upheld, for response tendencies to both cues appeared to adapt with long training series; and this result was the same whether or not *Ss* had been tested at the short training point.

Some verification of the definition of the feasibility dimension can be obtained from a tabulation of data from the postexperimental questionnaire. Of the 80 *Ss*, 67 reported that they expected to receive an electric shock following the verbal cue; whereas only 12 said that they expected to experience scalding water. When asked whether they actually experienced an electric shock, 12 said yes; whereas 1 of the 80 reported experiencing the scalding water.

The evidence supports the notion that perceptual sets in the form of tendencies to anticipate stimulation can be manipulated experimentally. By presenting verbal test cues during a period when other verbal cues are being reinforced, the capability of the test cues to elicit anticipatory GSRs is enhanced. The changes that occur in the response tendency (set) would appear to be first one of increase with associated reinforcement training, then a reduction (adaptation effect) accompanying further training.

It is proposed that the experimental situation discussed has usefulness in studies of the reinforcement process with human *Ss*, particularly where concern is for modification of sets to anticipate stimulation. The situation does not involve direct reinforcement but employs a kind of atmosphere effect wherein nondirect reinforcement can be influential in modifying a response set. It can be further emphasized that such nondirect reinforcement may be the only type that is effective in such a situation. If the noxious suggestion cues were reinforced directly, it is probable that response adaptation would begin immediately; whereas, by focusing on the anticipation process and confirming or reinforcing some anticipations, a phenomenon is observed which would not otherwise occur.

Summary

A set or tendency to anticipate that stimulation will follow a verbal cue suggesting such stimulation is inferred from the magnitude of GSR to the cue. The experimental procedures for modifying the set are analogous to classical conditioning procedures, except that the reinforcing state of affairs is not applied directly to the test cues but is applied instead to a series of seven different cues.

Experimental variation is introduced in a 2×2 factorial arrangement with one inde-

pendent variable the number of reinforcements applied before test, and the second independent variable the nature of the test cue. Half of the 80 *Ss* were tested on a cue involving a feasible suggestion of stimulation. The other half were tested on a cue designated unfeasible.

Plots of magnitude of response to the critical test cues as a function of position in the reinforcement series showed an increase of response to both cues at the short training test point, then a decrease with further training. The principal change occurred for the feasible stimulus during the short series of reinforcements.

REFERENCES

ANOKHIN, P. K. Characteristics of the afferent apparatus of a conditioned reflex and its importance for psychology. *Vop. Psikhol.*, 1955, No. 6, 16–38.

BRUNER, J. S. Personality dynamics and the process of perceiving. In R. R. Blake and G. V. Ramsey (Eds.), *Perception: An approach to personality.* New York: Ronald, 1951.

CORNBECKER, F., WELCH, L., & FISICHELLI, V. Conditioning factors underlying hypnosis. *J. abnorm. soc. Psychol.*, 1949, **44,** 212–222.

GRINGS, W. W. Preparatory set variables related to classical conditioning of autonomic responses. *Psychol. Rev.*, 1960, **67,** 243–252.

HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.

MOWRER, O. H. *Learning theory and behavior.* New York: Wiley, 1960.

RAZRAN, G. H. S. Conditioning and perception. *Psychol. Rev.*, 1955, **62,** 83–95.

SOLLEY, C. M., & MURPHY, G. *Development of the perceptual world.* New York: Basic Books, 1960.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

WOODWORTH, R. S. Re-enforcement of perception. *Amer. J. Psychol.*, 1947, **60,** 119–124.

WOODWORTH, R. S. *Dynamics of behavior.* New York: Holt, 1958.

(Received February 24, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 423-425

ACUITY AND THE STATISTICAL THEORY OF FIGURAL AFTEREFFECTS

F. H. GEORGE

University of Bristol

An experiment was carried out on visual acuity under conditions of fatigue similar to those met in the figural aftereffect situation. This experiment was performed with the object of testing a supposition made to preserve the general form of the Osgood-Heyer (1952) type of explanation against some objections made by Deutsch (1956). The supposition was that whereas in the ordinary course of events summation of excitation, to the extent suggested by Osgood and Heyer, does *not* occur in the visual cortex, under conditions of fatigue, such as in the inspection period of the figural aftereffect experiments, there is a spread of excitation that brings about summation over very considerable areas. This effect was thought to be more marked with respect to areas in the periphery of the retina. A simple way of testing this supposition offers itself in the use of a situation that demands discrimination of two contour lines under "fatiguing" conditions. The results obtained suggest possible modification to the Osgood-Heyer model.

The principal fault in the Osgood-Heyer (1952) model, according to Deutsch (1956), is that it uses an altogether too versatile concept of summation. The apparent shift of the contour line in the figural aftereffect situation depends upon the new contour line falling on the area already fatigued, the distributions thus overlapping and causing summation. But the argument goes on to say that these adjacent lines are easily discriminated within the area over which the figural aftereffect shift is supposed to take place. Now clearly Osgood and Heyer cannot have it both ways; the problem exists as to how we can discriminate lines at all, *within* the range of the shift, for either there is a failure of discrimination—where, incidentally, it is unquestionably known to be possible—because overlapping distributions summate with a resulting shift, or we can discriminate, in which case we do not get a shift.

METHOD

Subjects.—Twenty *S*s were used; they were all ordinary citizens of Bristol or undergraduates of the university. No prior check was made on their eyesight.

Apparatus.—The apparatus was made up of rectangular pieces of 8 × 12 in. white cardboard, on each of which were drawn pairs of vertical lines and fixation points.

Procedure.—The cards were presented to *S*s in a small stand and under standard conditions of illumination. There is little doubt that the results would be affected to some extent by the strength of the illumination, but this experiment was designed merely to discover whether the ordinary discrimination of lines would occur when slightly displaced from the fovea, and under conditions of "fatigue."

Three different pairs of adjacent lines were used at three different distances from a fixation point; there was one pair of lines on each card, each line being 2 in. long and $\frac{1}{16}$ in. thick. On the first card (A) the lines were $\frac{1}{16}$ in. apart, and the three fixation points were 2, 4, and 6 in., respectively, from the nearest of the pairs of lines. On Card B the two lines were $\frac{1}{8}$ in. apart, and the fixation points were the same distance away from the nearest line as on Card A. On Card C the lines were $\frac{3}{16}$ in. apart, and again the fixation points were the same distances from the nearest line as they were on Card A. The cards were

TABLE 1

NUMBER OF *Ss* SEEING ONE LINE, TWO LINES, OR NEITHER (INDEFINITE)

Card	Fixation Point	Binocular Viewing						Monocular Viewing					
		Lines to Right			Lines to Left			Lines to Right			Lines to Left		
		1 Line	2 Lines	Indef.	1 Line	2 Lines	Indef.	1 Line	2 Lines	Indef.	1 Line	2 Lines	Indef.
A	2 in.	9	0	1	10	0	0	8	0	2	9	0	1
	4 in.	8	1	1	8	1	1	8	1	1	10	0	0
	6 in.	9	0	1	8	0	2	9	0	1	7	2	1
B	2 in.	9	0	1	8	1	1	7	2	1	9	1	0
	4 in.	8	1	1	7	2	1	8	1	1	8	0	2
	6 in.	8	2	0	7	1	2	8	1	1	7	2	1
C	2 in.	4	3	3	2	7	1	1	7	2	2	7	1
	4 in.	1	5	4	0	8	2	0	10	0	0	9	1
	6 in.	0	8	2	0	7	3	0	9	1	0	7	3

all viewed from a distance of 18 in. from the midpoint between the eyes.

The *Ss* were instructed to fixate a particular fixation point, and to continue to fixate it for 1 min., at the end of which *E* said, "Keep fixating the fixation point carefully, and tell me how many lines you are aware of on the left, now." The word "right" was used in place of "left" when the lines were on the other side.

This same request was repeated after a further 30 sec. of fixating. The lines on the card were not made visible until some seconds after *S* had started fixating, so that he did not have an ordinary, clear, prior view of the pairs of lines.

The *Ss* were asked to view all three fixation points for each card, first with the lines on the left and then with the lines on the right. They were then asked to repeat the whole procedure monocularly, the verbal instructions being amended to read, "Close one eye and fixate the fixation point with the other eye." After 1 min. the instruction continued, "Now open the closed eye and shut the other one, and fixate the same fixation point, and tell me how many lines there are on the left (or right) of the fixation point."

As a precaution, the order of presentation was sometimes changed to monocular-binocular, as opposed to the above order, and within this the nine different cases were presented in a random order.

RESULTS AND DISCUSSION

Since the experiment was not really intended to be a highly refined one, but merely a test of an hypothesis of a straightforward kind—although requiring controlled conditions—the results were scored simply as one line or two or indefinite, according to *S*'s reply when questioned. Table 1 shows all the results in tabular form. It should be added that a control group of 20 *Ss* was used, and for all positions of the lines they were able to detect the presence of two lines when asked within 5 sec. of viewing the figures. This control group did not, of course, go through any fixation period (inspection period).

The results in Table 1 are clear cut in that a significantly large number of *Ss* under all the conditions saw the two lines disappear into one for all of cases of Card A and all of cases of Card B, but only a few for all of cases of Card C. These figures are, of course, highly significant. The rows marked "indefinite" simply contain the results of those *Ss* who were unable to say whether they saw one or two lines. But first it must be remarked that many *Ss* recorded many other observations of interest, although for the most part familiar to workers in the field of the figural

aftereffect, and these cannot be discussed here.

The results offer confirmation of the argument that the effect of fatiguing the eye by steady fixation is to spread the distribution of excitation in such a way as to cause summation to occur much more readily in the visual cortex. The fact that it was nearly as effective monocularly, when the eyes were exchanged before the decision was taken, shows that the effect is almost certainly central.

It should be mentioned that many more *Ss* were able to see the effect informally; in fact, anyone may draw two such lines, with a dot for fixation purposes, and convince himself of the facts.

This all means that the principal objection to the Osgood-Heyer model has been circumvented, and that it is perfectly possible to account for both figural aftereffects and ordinary visual discrimination, still within the ordinary interpretation of summation.

Nevertheless, some amendments do have to be made to the Osgood-Heyer model. They are to the effect that the spread of distribution is a function of the fatiguing, and not merely that with fatigue there is a diminution of excitation. Similarly, but for other reasons, they may have to drop the idea that eye movements are *necessary to* the original distribution of excitation (Deutsch, 1956; Riggs, Ratliff, Cornsweet, & Cornsweet, 1953). But there is now some additional, although perhaps minor, reason for believing this to be the case, and that is that we may expect a considerable disparity between the girth of the distribution of excitation in the ordinary course of a line being projected onto the retina, and the case where the line is projected onto the retina when fixation of some other point occurs for a period of 1 min. or more. The difference here could be between the narrow ridge of excitation caused by the line, and the spread of excitation caused by fatigue. No attempt will be made here to relate the assumed spread of excitation with fatigue to the change of activity in the retinal elements.

Summary

This experiment was an attempt to reinstate the Osgood-Heyer (1952) model for explanatory purposes in perception, with the object of making plausible extension of that model in order to explain movement aftereffects.

The experiment involved sets of double-line figures at various distances from a fixation point and *Ss* were unable to distinguish the two lines when sufficiently far from the fixation point, *and* under conditions of "fatigue." Thus, the principal features of the Osgood-Heyer model can easily be preserved from the main objections made against them by Deutsch (1956).

References

Deutsch, J. A. The statistical theory of figural aftereffects and acuity. *Brit. J. Psychol.*, 1956, **47**, 208–215.

Osgood, C. E., & Heyer, A. W. A new interpretation of figural after-effects. *Psychol. Rev.*, 1952, **59**, 98–118.

Riggs, L. A., Ratliff, F., Cornsweet, J. C., & Cornsweet, T. N. Disappearance of a steadily fixated visual test object. *J. Opt. Soc. Amer.*, 1953, **43**, 495–501.

(Received January 31, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 426-437

ASSIMILATION VS. CONTRAST IN THE ANCHORING OF PERCEPTUAL JUDGMENTS OF WEIGHT [1]

ALLEN PARDUCCI AND LOUISE M. MARSHALL

University of California, Los Angeles

Successive perceptual contrast may be demonstrated for comparative judgments by interpolating an extra stimulus between the standard and comparison stimuli. The value of the comparison judged equal to the standard, the PSE, varies directly (and the judgments inversely) with the value of the interpolated stimulus (Guilford, 1954). Successive contrast can also be demonstrated in the absence of a specified standard, i.e., for absolute judgments. When a new stimulus value is added to a series presented for judgment by the method of single stimuli, the mean judgment of the remaining stimuli varies inversely with the value of the new stimulus. This contrast occurs both when the end-category of the absolute scale is anchored to the new stimulus (e.g., Rogers, 1941; Sherif, Taub, & Hovland, 1958) and also when the new stimulus is presented without specific anchoring instructions (Parducci, 1954; Postman & Miller, 1945).

These examples of perceptual contrast are quantitatively described by the theory of adaptation level (Helson, 1947, 1959). According to this theory, the PSE or the adaptation level, AL, the stimulus value at which the scale of judgment is centered, corresponds to a weighted logarithmic mean of the various physical values of the stimuli. Each of the judgments is in turn determined by the ratio of the stimulus being judged to AL. To the degree that the introduction of a new stimulus shifts the mean (and thus AL) toward its own physical value, contrast or an opposing shift in judgment occurs.

However, the results of a recent experiment with lifted weights have raised doubts concerning the generality of this description of judgment (Sherif et al., 1958). Although the expected contrast effects were found for anchors outside the range of the regular series, the opposite effect (i.e., heavier judgments of the series weights) was produced by an anchor equal in weight to the heaviest member of the regular series. Sherif et al. (1958) applied the term *assimilation* to this phenomenon, interpreting it as an analogue to certain propaganda effects of social communication. They also noted that such assimilation is not easily explained in terms of the adaptation-level approach.

The present experiments were designed to evaluate this apparently negative instance with respect to an alternative application of the theory of adaptation level. Instead of assuming that the anchoring stimulus serves as just another member of the series presented for absolute judgment, it is here proposed that the anchor should be treated as a standard for comparative judgment. Since the experimental instructions used by Sherif et al. (1958) state that the anchoring stimulus is to be judged with the heaviest category, PSE is by definition the physical value of the stimulus actually judged with the heaviest category. According to the theory of adaptation level, PSE is

[1] This research was supported in part by a grant (M-3134 A) from the National Institutes of Health.

equal to a weighted logarithmic mean of the standard and comparison stimuli (Helson, 1959; Michels & Helson, 1954). When the standard is identical to the heaviest comparison weight, PSE will be below the value of the standard since any weighted mean of the stimulus values would have to be lower than the standard. But if PSE is below the actual physical value of this anchoring stimulus, the heaviest comparison weight is perceived to be heavier than the anchor—perceptual contrast. The anchoring conditions employed by Sherif et al. (1958) thus lower the value of the stimulus to which the "heaviest" category of judgment is applied, and the entire scale of judgment is shifted downward.

The present hypothesis is that this downward shift, the assimilation found by Sherif et al. (1958), is a special case of the *central-tendency effect* (Hollingworth, 1910; Long, 1937). As described by the theory of adaptation level, the central-tendency effect is basically a phenomenon of perceptual contrast. The contrast lies in the shift in PSE toward a weighted mean of the experimental stimuli. The following experiments evaluate this interpretation by varying the instructions for judgment.

Experiment I: Test for Central Tendency

The first objective was to replicate the relevant portions of the experiment reported by Sherif et al. (1958) with additional controls for the order of presentation of the respective experimental conditions. To test the present interpretation of assimilation, two additional conditions were introduced which permitted direct assessment of the PSE for the anchoring stimulus. For one condition, a different category of judgment was fixed to the anchoring stimulus; for the other, an additional category was added to the scale of judgment.

Method

Stimuli.—The six weights used as the "original stimulus series" by Sherif et al. (55, 75, 93, 109, 125, and 141 gm.) were also used for every condition of the present experiments. The individual weights were presented for judgment by means of an apparatus constructed in accordance with the description reported for the previous experiment. Essentially, it served to block *S*'s view of the weights and of *E*, the lifting being accomplished by pulling down a handle which was attached to a cord leading over a pair of low-friction wheels to the weight.

Procedure.—The instructions described by Sherif et al. (1958) were used for three of the present conditions. Thus for Cond. No-A (No-Anchor), *Ss* had to make their judgments in terms of an absolute scale of 1 to 6, inclusive, with no special anchoring of any of the response categories (i.e., *Ss* were not told the number which went with each weight, or even how many weights there were). For Cond. H-A (Heavy-Anchor), *Ss* were told that the stimuli would be presented in pairs, the first member of each pair indicating the weight which they were to call 6. This was the 141-gm. weight. The 55-gm. weight was presented as the anchor for Cond. L-A (Light-Anchor), and *Ss* were instructed to call it 1.

The 141-gm. weight was used as the anchor for both of the new conditions. It was again labeled 6 for Cond. 7-C (7-Category), but *Ss* were instructed to make their judgments in terms of a 7- rather than a 6-point scale. The anchor was labeled 5 on a 6-point scale for Cond. Anchor-5. These conditions can perhaps be more readily distinguished by reference to Fig. 1, with special attention to the location of the different anchoring weights with respect to the response scale.

Each condition required a total of 288 judgments by each *S*. In addition, there were 288 unjudged presentations (the odd trials) for each of the anchor conditions. The order of presentation was randomized over blocks of 48 trials (or pairs of trials when an anchor was used) so that each of the six weights appeared eight times within each block. An experimental session lasted somewhat under 1 hr., with two short rests permitted within the session. Each *S* served in a

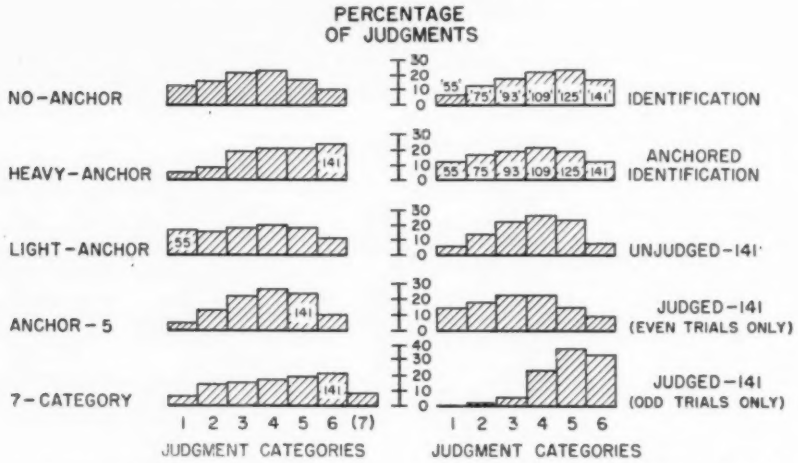

FIG. 1. Distribution of judgments for conditions selected from Exp. I, III, and IV. (Anchored categories indicated by weight, gm., of anchoring stimulus.)

minimum of two experimental sessions which were always on different days.

Subjects and order of the conditions.—Each of the groups serving in these experiments was composed of 10 *Ss* drawn from the introductory psychology course at the University of California, Los Angeles. Following the procedure reported by Sherif et al. (1958), the two replication groups served under more than one of the conditions being compared.

TABLE 1

MEAN ADAPTATION LEVELS AND PERCENTAGE DISTRIBUTION OF JUDGMENTS

Condition	Categories of Judgment						AL	*SD*
	1	2	3	4	5	6		
No-Anchor (1st)[a]	13	16	21	23	17	10	103	5.8
Heavy-Anchor (1st)	5	9	20	21	21	24	87	11.4
No-Anchor (2nd)	6	13	21	24	20	16	94	7.1
Heavy-Anchor (2nd)	11	14	16	17	21	21	93	8.6
Anchor-5	5	13	22	27	23	10	94	6.7
Anchor-5 (3rd)	14	16	18	23	23	6	101	6.9
Light-Anchor	17	16	18	20	18	11	105	10.1
7-Category[b]	6	14	15	17	19	21	95	6.9
Heavy-Standard	15	24	33	22	5	1	124[c]	11.5
Light-Standard	0	2	11	31	33	23	56[c]	14.8
Unjudged-141	6	14	22	27	23	8	95	10.6
Judged-141	14	18	22	22	15	9	105	5.9
Judged-141[d]	7	9	14	22	27	21	114	8.3
Identification	7	13	18	22	23	17	93	9.2
Anch.-Identification	12	17	19	21	19	12	101	2.5

[a] Order of presentation indicated only when *Ss* judged under more than one condition.
[b] Category 7 was used for 8% of the judgments.
[c] PSE for the standard.
[d] Including judgments of 141-gm. weight on odd trials.

However, the order of the conditions was systematically varied in the present design so that the effects of order could themselves be assessed. Thus one group was exposed first to Cond. No-A, then to Cond. H-A, and finally to a new condition, Anchor-5. A second group was exposed to the first two of these conditions only; for this group, Cond. H-A preceded Cond. No-A. A third group served only under Cond. Anchor-5. Each of the other groups in these experiments also served under only one of the experimental conditions.

Results and Discussion

The percentage of the 288 judgments made in terms of each of the respective categories was computed for each *S* under each condition. Table 1 shows these percentages averaged over 10 *S*s. To facilitate comparison with the figures presented in Sherif et al. (1958), each group's distribution of judgments for the condition it judged first is presented graphically in Fig. 1. As compared with Cond. No-A, composed of the same group of *S*s, Cond. H-A produces a marked assimilation effect, i.e., more frequent use of the heavier categories of judgment for H-A (2nd) than for No-A (1st), for H-A (1st) than for No-A (2nd). To evaluate statistically this assimilation, the percentage of use of each category by each *S* under Cond. H-A was subtracted from his corresponding percentage under Cond. No-A. These percentage differences were analyzed according to a two-factor, Judgment × Order design, different *S*s being used for the two orders but the same *S* having a difference score for each of the six categories. Since the sum of the six difference scores was automatically zero for each *S*, the effects of Order and *S*s were necessarily zero. The Judgment *MS* was highly significant in this analysis ($F = 11.71$, $df = 5/90$, $P < .001$, using as the error term the pooled interaction between *S*s and Judgment—as described in Lindquist, 1953, pp. 267–272). The interaction between Order and Judgment was not significant ($P > .25$, again using the pooled *S*s × Judgment interaction as the error term). This means that the assimilation reported by Sherif et al. (1958) was found to be a reliable effect of the present replication (the distributions of judgments being significantly different for Cond. No-A and H-A). It also means that there is scant evidence of the magnitude of the assimilation being affected by the order in which the two conditions are presented.

However, the distributions of judgments under Cond. No-A, H-A, and Anchor-5, respectively, are each affected by whether or not the given condition has been preceded by other conditions, i.e., whether a given condition was the first, second, or third of those judged by the 10 *S*s following each of the two orders of conditions. Judgment × Order analyses of variance were performed for each of the following three pairs of conditions: No-A (1st) vs. No-A (2nd), H-A (1st) vs. H-A (2nd), and Anchor-5 vs. Anchor-5 (3rd). The percentage-use of each of the categories by each *S* was used for the raw scores. The interaction between Order and Judgment was significant in each of the three analyses ($P < .005$, .025, and .01, respectively). Thus while the present replication confirms the assimilation to the heavy anchor reported by Sherif et al. (1958), it also suggests that their procedure of exposing the same group of *S*s to the various experimental conditions markedly affects the distribution of judgments under any given condition. Examination of the direction of this effect indicates that, in each case, the distribution of judgments tends to resemble the distribution of judgments for the previously presented condition(s). This appears consistent with an adaptation-level description since the respective ALs are pulled toward the value of preceding

ALs. The ALs, defined as one-half the sum of the means of the stimuli judged 3 and 4, respectively, are presented in Table 1.

An analysis of variance based on the percentage scores for Cond. No-A (1st) and Cond. L-A indicated that the difference between these two distributions of judgments was not significant ($F = .90$). This absence of assimilation appears inconsistent with the general trend of the light-anchor effects reported by Sherif et al. (1958). However, they introduced the following experimental variations for this comparison: direct lifting of the weights, fewer and heavier weights, fewer categories, and fewer *S*s. While their own data are not entirely consistent for the different light-anchor conditions they report, there is some indication that they obtained less assimilation to light than to heavy anchors. This suggests that a negative time-order error may be counteracting the postulated central-tendency effect for the light-anchor condition. Negative time-order errors are usually found for lifted weights when the comparison stimuli follow the standard, i.e., the comparison weights tend to be judged heavier than the standard, even though they are physically equal to the standard in weight (Michels & Helson, 1954). Additional evidence, consistent with the assumption that this type of constant error is operating in the present situation, will be presented below under Exp. II.

To test the central-tendency interpretation of the assimilation to the heavy anchor, the PSE for the 141-gm. anchor (i.e., the weight judged equal to the anchor) was determined for both Cond. Anchor-5 and Cond. 7-C by calculating the mean of the upper and lower limens for the anchored category. This could not be done for Cond. H-A since the anchored category (i.e., 6) had no upper limen. The respective limens were first determined by a least squares fit to the z scores corresponding to the percentage of times each stimulus was judged above the limen in question (Guilford, 1954, pp. 125–128). The resulting Anchor-5 and 7-Category PSEs were 132 and 135 gm., respectively. Since the anchor was 141 gm. in both cases, the data are consistent with the assumption that there was a negative constant error for this anchor (i.e., in the direction of the mean of the experimental stimuli).

Experiment II: Comparative Judgment

While the anchoring instructions used in Exp. I tell *S*s how they are to judge weights which appear equal to the anchoring weight, it is possible that these instructions conflict with the more general instructions for absolute judgment. Insofar as such conflict occurs, the PSEs determined for the Anchor-5 and the 7-Category conditions may not indicate the weight which *S*s would actually perceive as equal to the anchor. In the present experiment, the method of constant stimulus differences is used to obtain a more direct measure of the PSE for the two anchor weights used in Exp. I.

Method

The instructions stated that *S* would be presented weights in pairs and that his task was to judge how heavy the second member of each pair felt in comparison with the first, the standard. Judgments were made in terms of the following categories: much lighter—1, lighter—2, slightly lighter—3, slightly heavier—4, heavier—5, and much heavier—6. These comparative categories yield convenient psychometric functions (Michels & Helson, 1954), and their use in place of the usual restriction to two or three categories increased the similarity between Exp. I and II. In other respects, the former procedure was followed: the same six weights, the same randomization of the comparison presentations in blocks of 48, and the same experimental apparatus. Two new groups of 10 *S*s each were run, one under Cond. H-S (Heavy-Standard, 141 gm.) and the other under Cond. L-S (Light-Standard, 55 gm.).

Results and Discussion

A PSE was obtained for each *S* by tabulating the proportion of his judgments of each stimulus which were either 4, 5, or 6 (i.e., judged "heavier" than the standard), converting these proportions to z scores, and then calculating his PSE by the method of least squares. The mean and *SD* of Cond. H-S PSEs were 124 and 12 gm., respectively, with 9 of the 10 PSEs lower than the 141-gm. standard. The mean of these constant errors is significantly negative ($t = 4.43$, $df = 9$, $P < .005$).

Since 6 of the Cond. L-S *S*s judged the lightest comparison weight to be heavier than the standard (both were 55 gm.) on over half the presentations of this pair and since 4 of these 6 judged every presentation of the remaining comparison stimuli heavier than the standard, it was not appropriate to use the same procedure for determining Cond. L-S PSEs. Linear interpolation (or extrapolation) for the mean judgments of the respective stimuli was used instead. This procedure yielded almost the same PSE for Cond. H-S, 123 gm., and a Cond. L-S PSE of 56 gm. (as shown in Fig. 2 where the two PSEs correspond to mean judgments of 3.5 on the respective psychometric functions). The difference in magnitude between the two constant errors (−18 gm. vs. +1 gm.) is consistent with the assumption, made in the previous discussion, that a negative time-order error was counteracting the effects of the light anchor in Exp. I. Thus the constant errors obtained when the stimuli are explicitly compared with the anchoring weights (the standard for the method of constant stimulus differences) correspond to the shifts in the scales of judgment obtained when the absolute judgments are preceded by these same anchoring weights.

It should be noted that the Cond. H-S PSE is lower than the PSEs obtained for this same weight under the Anchor-5 and 7-Category conditions of Exp. I. This difference may be due to differences between the absolute and comparative response scales. As shown in Table 1, the distributions of judge ments for the comparative conditions are much less symmetrical (e.g., very few 5 and 6 judgments for Cond. H-S) than the absolute scales for the same stimuli (e.g., for Cond. Anchor-5). This is a consequence of the difference in the instructions for judging the anchoring weight and the standard, the 141-gm. weight being labeled either 5 or 6 in the anchoring instructions used in Exp. I but only 3.5 in terms of the categories for comparative judgment used in Exp. II. However, the difference in symmetry due to this difference in the instructions for judgment would be reduced by the tendency for *S*s to use their respective categories with fixed proportionate frequencies, symmetrical about the middle of the response scale (i.e., 4, 5, and 6 as often as 1, 2, and 3—as asserted with

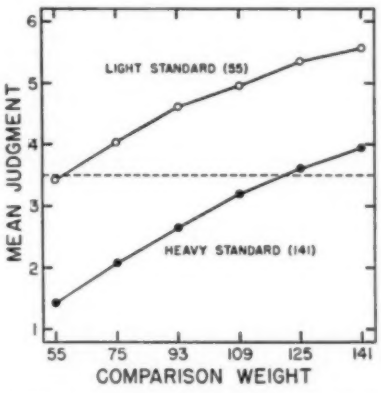

FIG. 2. Mean judgments of each of the comparison weights following either the light or heavy standards of Exp. II. (All judgments were obtained by the method of constant stimulus differences.)

respect to the median-midpoint compromise demonstrated in Parducci, Calfee, Marshall, & Davidson, 1960). The compromise toward a symmetrical scale of judgment would thus produce a lower PSE for Cond. H-S than for either the Anchor-5 or 7-Category conditions. But whatever the value of this interpretation of the differences between the PSEs, it is clear that the PSEs obtained under the comparative instructions are consistent with the adaptation-level interpretation of assimilation.

Experiment III: Unlabeled Anchors

The adaptation-level interpretation of assimilation stresses the anchoring instructions provided by *E*. Without verbal anchoring, the presentation of the 141-gm. weight on alternate trials should produce contrast rather than assimilation (Helson, 1947). Since this extra weight would raise the mean of the stimuli, it would also raise AL and hence lower the judgments.

Method

To test this implication of adaptation-level theory, two additional groups of 10 *Ss* each were presented the same series of 576 weights as in Cond. H-A of Exp. I. For Cond. Unjudged-141, *Ss* were instructed not to judge the weight presented on every other trial (i.e., the odd-numbered trials); for Cond. Judged-141, *Ss* made a judgment on every trial—576 judgments, with no special attention drawn to the alternate trials. Neither group was given anchoring instructions.

Results and Discussion

As may be inferred from comparison with the distribution of judgments for Cond. No-A (see Table 1 and Fig. 1), assimilation rather than the predicted contrast was obtained under Cond. Unjudged-141. The Condition × Judgment × *Ss* analysis of variance indicated that the distribution of judgment frequencies for this condition differed significantly from the distribution for Cond. No-A (1st) ($F = 3.98$, $df = 5/90$, $P < .005$). There is a striking correspondence, however, between the Unjudged-141 and the Anchor-5 distributions. Since tabulation of the 288 odd-trial judgments for Cond. Judged-141 (see Fig. 1) indicates that the 141-gm. weight was judged below 5 almost as often as it was judged 6 (below 5–29%; 5–38%; 6–33%), it appears that the Unjudged-141 *Ss* may have anchored their own scales by implicitly labeling each interpolated (i.e., odd-trial) weight as 5 on their 6-point scale.

Why, then, should Cond. Judged-141 yield such a different distribution of judgments (for Judgment × Condition interaction, Judged-141 vs. Unjudged-141, $F = 4.01$, $df = 5/90$, $P < .005$)? Again it seems appropriate to appeal to the postulated tendency to use the categories with symmetrically proportionate frequencies (Parducci et al., 1960). Table 1 shows two different distributions of judgments for Cond. Judged-141. The first is based upon the even trials only (i.e., on 288 judgments); the second is based upon all 576 trials (including the judgments of the odd-trial presentations of the 141-gm. weight). The second is naturally much more asymmetrical (negatively skewed) than either the first or the distribution for Cond. Unjudged-141 since it includes the judgments of 288 extra presentations of the 141-gm. weight (shown in Fig. 1). The hypothesis here is that the difference between the even-trial distributions for Unjudged-141 and Judged-141 reflects the tendency, under the latter condition, to reduce the negative skewing of the total distribution of judgments.

Experiment IV: Identification

Experiments using absolute judgments have sometimes specified that *Ss* were to use their categories of judgment to identify or discriminate

the alternative stimuli. Such judgments are *absolute* in the sense that a judgment is made of each stimulus and the implied standards for judgment are not identified during the series of presentations (i.e., the procedure uses the method of single stimuli rather than the method of constant stimulus differences). However, since *S* is told that each category of judgment corresponds to (or identifies) a different stimulus value, each judgment can be scored as either correct or incorrect.

In Exp. I and III, there were six weights and also six categories (except for the 7-Category condition); but this correspondence was never stated explicitly, *S*s being told only that they should judge each weight on a scale from 1 to 6. The difficulty of discrimination was such that *S*s could not learn that there were exactly six weights, and they may not have assumed a one-to-one correspondence between stimuli and categories. Thus the task for judgment in these experiments may resemble closely the task for absolute judgment as established in those experiments for which the number of categories does not correspond to the number of stimuli (e.g., Helson, 1947; Parducci, 1954; Parducci et al., 1960). Such absolute judgments can be neither correct nor incorrect, and the explicit instructions require only that the judgments reflect the psychological relationships among the stimuli. However, the results of recent research have suggested that the use of absolute categories is further restricted by implicit rules of judgment. Thus there is a tendency to use the categories to divide the range of stimuli into proportionate subranges and also to use the categories with fixed relative frequencies (Parducci et al., 1960). Insofar as these implicit rules of judgment do not reflect an attempt simply to identify or discriminate between the alternative stimuli, Exp. I and III might have yielded quite different distributions of judgment if *S*s had been explicitly instructed to identify the stimuli.

Method

To compare the effects of specific instructions to identify the stimuli with the effects of the task for absolute judgment as employed in Exp. I and III, two new groups of 10 *S*s each were told that there were six different weights which they were to identify with the numerals 1 through 6, respectively. Before making any identifications, they were permitted to lift each of the six weights, in random order, without being told which weight was which. The procedure for the Cond. I (Identification) followed the Cond. No-A procedure in all other respects. The procedure for Cond. A-I (Anchored-Identification) was the same as for Cond. H-A except that all six weights served as anchors. Each of the judged weights followed each of the anchors within each successive block of 36 paired presentations. The anchoring weights were correctly identified by *E* on the odd trials, *S*'s task being to identify the weights presented on the even trials.

Results and Discussion

As shown in Table 1 and Fig. 1, the response distribution is more negatively skewed for Cond. I than for Cond. No-A ($F = 6.31, df = 5/90$, $P < .001$, for the Condition × Judgment interaction). The only difference betweeen these two conditions is in the instructions, i.e., whether *S* was to identify each stimulus with its own numeral (from 1 through 6) or whether he was to judge the stimuli on an absolute numerical scale (also from 1 through 6). The significant difference between the resulting response distributions thus suggests that the absolute scale does not simply reflect *S*'s attempts to identify the stimuli, even when he

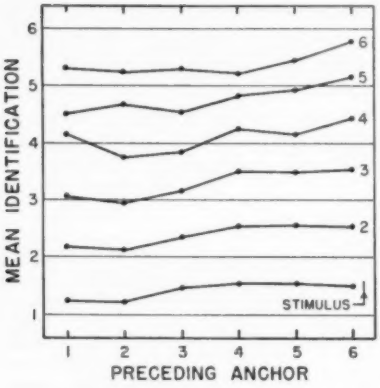

FIG. 3. Mean identification as a function of the preceding anchor for Anchored-Identification condition of Exp. IV.

uses the same number of response categories as there are stimuli.

Paradoxically, *Ss* were actually more accurate in identifying the respective stimuli when the correct response for each stimulus was not specified in the instructions (i.e., under Cond. No-A of Exp. I) than under the instructions for identifications (with no interpolated anchoring). The mean percentage of correct judgments was 44% for Cond. No-A (1st), but only 36% for Cond. I ($t = 2.02$, $df = 18$, $P = .05$). The greater accuracy following the Cond. No-A instructions is consistent with the assumption of a tendency to use the categories of judgment with fixed relative frequencies, symmetrical about the middle limen (Parducci et al., 1960). This hypothetical tendency is reflected in the striking symmetry of the Cond. No-A (1st) distribution of judgments as compared with the marked negative skewing found for Cond. I. The skewing of the identifications is presumably due to the negative time-order error demonstrated in Exp. I and II. Thus insofar as the PSEs for the heavier weights are below their actual physical values, *Ss* identify them as heavier than they actually are. This is another application of the concept of self-anchoring which was used to explain the assimilation found in Exp. III. The time-order error has less effect following the instructions for Cond. No-A since they place less restriction on *S*'s tendency to distribute his categories with proportionate frequencies.

As might be expected on the basis of the greater amount of information given, accuracy was greater for Cond. A-I than for Cond. No-A (52% vs. 44% correct, $t = 2.37$, $df = 18$, $P < .05$). Separate tabulation was made of the identifications of each stimulus following each anchor for Cond. A-I, and the mean identification of each stimulus is plotted against the value of the preceding anchor in Fig. 3. The slopes of all six of the resulting curves are positive. A trend test, based upon the linear components of these curves, indicated that while the slopes did not differ significantly from each other ($F = 1.57$, $df = 5/45$, $P > .10$, for Stimuli tested against the *Ss* × Stimuli interaction), the mean slope

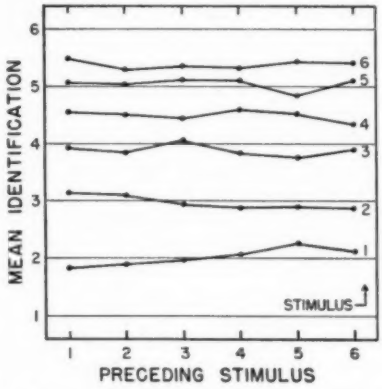

FIG. 4. Mean identification as a function of the preceding stimulus for Identification condition of Exp. IV.

was significantly greater than zero ($F = 24.82$, $df = 1/9$, $P < .001$, for Mean tested against *Ss*). It may thus be concluded that the judgment of each stimulus varies *directly* with the value of the preceding anchor. This further case of assimilation is also consistent with the central-tendency interpretation, assuming that the constant errors for the anchor PSEs are more negative for the heavier than for the lighter anchors and that the most recently presented anchor has a greater effect upon judgment than the preceding anchors.

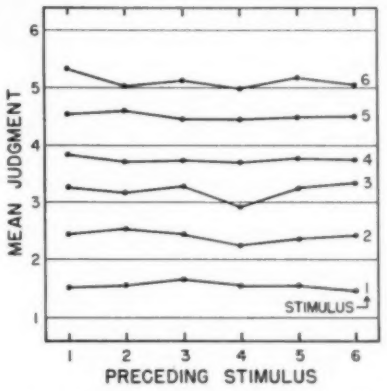

FIG. 5. Mean judgment as a function of the preceding stimulus for No-Anchor condition of Exp. I.

Similar curves have been reported for identification-judgments of loudness made without specific anchoring of the stimuli except during an initial practice session (Garner, 1953). It may be that Garner's situation produced considerable self-anchoring. However, the same tabulation for the present Cond. I and No-A (using, as Garner did, the preceding stimulus value instead of the preceding anchor) revealed no systematic variation in response as a function of the preceding stimulus (see Fig. 4 and 5). Garner proposed two alternative interpretations for his assimilation data. The first was that *Ss* confuse the presented stimulus with the preceding one. This would not explain why assimilation was obtained for the present identifications only when they were anchored. Garner's second interpretation was that *Ss* tend, when in doubt, to repeat their preceding identifications. The actual mean repetition rate for the unanchored identifications in the present experiment is only 17.8%, not significantly greater than the actual repetition of the stimuli. However, the mean rate of repetition of the preceding anchor category under Cond. A-I is 20.6%, significantly greater than chance ($t = 2.71$, $df = 9$, $P < .025$). Thus the occurrence of assimilation is associated with a tendency toward repetition. A breakdown by category of the Cond. A-I repetitions indicates that the rate of repetition varies systematically with the categories. Insofar as the PSEs for the more extreme anchors are shifted further toward the mean of the stimuli (time-order error being removed by the multiple anchoring), repetition should be greater for the more extreme categories. The mean percentage of the total use of each category represented by repetitions is as follows: 24, 18, 19, 19, 20, and 29% for 1 through 6, respectively. Although statistical evaluation of these differences is precluded by the variation in the denominators (the actual frequency of use of the respective categories), the curvilinear form of the repetition function is consistent with the hypothesized central-tendency effect.

DISCUSSION

These experiments demonstrate that assimilation effects are found for a variety of simple judgmental tasks. However, in each of the present cases, the assimilation is consistent with the contrast principles described by the theory of adaptation level. Whenever a category of judgment is anchored to a specific stimulus, either by *E* or by *S*, the anchored category is applied to the PSE for the anchoring stimulus in accordance with the adaptation-level analysis for comparative judgment. The remaining judgments are largely deter-

mined by the physical ratios of each of the various stimulus values to this PSE.

A major ambiguity lies in the specification of the conditions under which *S* anchors his own scale of judgment. Self-anchoring was invoked to explain the assimilation found in Exp. III where a single, unlabeled, unjudged stimulus was interpolated on alternate trials into the regular series of stimuli (i.e., Unjudged-141). While such invocation must be somewhat arbitrary until the conditions for self-anchoring are more adequately studied, it is consistent with the close correspondence between the "self-anchored" scale of judgment and the scale anchored with the category which *S*s would otherwise apply to the anchoring stimulus (i.e., Anchor-5). The adaptation-level approach, even with this ambiguity, would appear to provide a more powerful analysis than the common use of assimilation and contrast as two, basically opposed principles of judgment, capable between them of describing any possible shift in judgment.

While the results of these four experiments provide a comparison between various kinds of absolute and comparative judgments, the analysis stresses the anchoring features common to all of the procedures studied. The net effect is to emphasize the role of comparative judgment, comparison between the judged stimulus and the values of previously presented standards or anchors. This suggests that instead of the common distinction between absolute and comparative judgments, it would be more appropriate to classify the various procedures for judgment with respect to the manner and degree to which the rules for judgment have been specified.

Unanchored absolute judgments, elicited by the customary instructions to use verbal categories (e.g., light, medium, heavy), would appear to be the least specified, the least anchored by *E*. The numerical scale used for Cond. No-A of Exp. I approximates this lowest degree of specification, and the distribution of judgments under this condition is similar to the distribution obtained using verbal categories, i.e., symmetrical for a rectangular distribution of stimuli, with more frequent use of the middle categories (see Parducci et al., 1960, p. 71). But even though the instructions for such scales involve little explicit anchoring, the categories of judgment are to a considerable extent anchored by implicit rules for judgment. Thus each of the No-Anchor *S*s used all six categories; and as indicated in Fig. 5, the order of the categories corresponds to the order of the stimuli. As has been shown in previous experiments, the addition of an unlabeled weight which was heavier than 141 gm. would have anchored Category 6 to the new value, producing a contrasting shift in the judgments of the series stimuli (e.g., Postman & Miller, 1945). Much of the research on absolute judgment has been directed toward the discovery of such implicit rules of judgment.

When individual stimulus values are not singled out for special attention, *S*s have greater difficulty in applying either the implicit rules of judgment or the anchoring rules provided by *E*. Thus identifications were actually less accurate following the instructions stating the one-to-one correspondence between categories and stimuli than following the No-Anchor instructions. The Identification *S*s had no constant physical landmarks to which they could anchor their categories, and it appeared that the instructions to identify conflicted with the implicit rule to use the categories with fixed relative frequencies.

The introduction by *E* of a constant physical landmark, such as the unlabeled anchor interpolated into the series for Cond. Unjudged-141 of Exp. III, provides a further degree of specification. Insofar as this stimulus is recognized as unchanging, the category which *S* applies to it is anchored at a fixed value of the stimulus. Explicit anchoring of this stimulus by *E*, as in any of the anchoring conditions of Exp. I, would completely determine the location of at least one point on the scale of judgment if it were not for the changes in PSE

for the anchored value. Similar constant errors (systematic shifts in PSE) were found following both the absolute and comparative instructions, though the shift in PSE for the 141-gm. weight was actually less using the absolute procedure. Again, the implicit rule to distribute the categories with proportionate frequencies may have produced the greater accuracy of the absolute scale.

The Anchored-Identification condition involved the most complete specification of the various categories of judgment. Although the generally negative time-order error was effectively eliminated under this condition, the assimilation toward the values of the preceding anchors (shown in Fig. 3) appeared consistent with the adaptation-level analysis of the constant errors in the PSEs for the anchored values. Thus the same sort of perceptual contrast was obtained when all six categories were repeatedly anchored to their respective stimuli as when there was less specification of the rules for judgment. The conclusion from this analysis is that while the nature and degree of specification determine what shifts in judgment will occur, the judgments are determined by the physical relationships between the judged stimuli and the PSEs for the anchoring or standard stimuli.

Summary

A successful replication was described for a recent demonstration of assimilation, rather than the usual contrast, in the anchoring of lifted weights. This assimilation was interpreted as a special case of the central-tendency effect, with the anchor serving as the standard for comparative judgments. Using both absolute and comparative judgments, it was shown that the PSE for the anchor shifted toward the center of the stimulus series in accordance with the contrast effects described by the theory of adaptation level. While it was necessary to invoke self-anchoring to explain assimilation to an unlabeled stimulus, this permitted reduction of all cases of assimilation to simple perceptual contrast. The analysis suggested that various procedures for absolute and comparative judgment are distinguished primarily on the basis of the nature and degree of specification of the rules for judgment.

References

Garner, W. R. An informational analysis of absolute judgments of loudness. *J. exp. Psychol.*, 1953, **46,** 373–380.

Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

Helson, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60,** 1–29.

Helson, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of science.* Vol. 1. New York: McGraw-Hill, 1959.

Hollingworth, H. L. The central tendency of judgment. *J. Phil. Psychol., sci. Meth.*, 1910, **7,** 461–468.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Long, L. A study of the effect of preceding stimuli upon the judgment of auditory intensities. *Arch. Psychol., N. Y.*, 1937, **30,** No. 209.

Michels, W. C., & Helson, H. A quantitative theory of time-order effects. *Amer. J. Psychol.*, 1954, **67,** 327–334.

Parducci, A. Learning variables in the judgment of single stimuli. *J. exp. Psychol.*, 1954, **48,** 24–30.

Parducci, A., Calfee, R. C., Marshall, L. M., & Davidson, L. P. Context effects in judgment: Adaptation level as a function of the mean, midpoint, and median of the stimuli. *J. exp. Psychol.*, 1960, **60,** 65–77.

Postman, L. J., & Miller, G. A. Anchoring of temporal judgments. *Amer. J. Psychol.*, 1945, **58,** 43–53.

Rogers, S. The anchoring of absolute judgments. *Arch. Psychol., N. Y.*, 1941, **37,** No. 261.

Sherif, M., Taub, D., & Hovland, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *J. exp. Psychol.*, 1958, **55,** 150–155.

(Received February 3, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 438-443

CUE NEUTRALIZATION: THE EFFECTS OF RANDOM REINFORCEMENTS UPON DISCRIMINATION LEARNING[1]

MARVIN LEVINE

Indiana University

The present experiments are concerned with the effects of random reinforcements (RRs) upon subsequent discrimination learning. Human *S*s are presented with a sequence of two colored stimuli (e.g., blue or tan cards), are instructed to make one of two responses to each stimulus, and are told either "right" or "wrong" following each response. During a first procedural phase, for a given number of trials, *E* applies the reinforcing events on a prearranged random schedule, i.e., the words "right" and "wrong" are uncorrelated with any stimulus-response pairs. During a second phase, which appears without special announcement or break, the traditional discrimination procedure is employed in which the word "right" is presented only when the correct response is made to the appropriate color. Otherwise the word "wrong" is presented. One may then inquire into the effect of the preliminary RRs upon the subsequent discrimination learning by evaluating the difficulty encountered in learning the discrimination problem.

The experimental results will have implications for the process of cue neutralization, and for theories which describe this process. Restle (1955; Bourne & Restle, 1959) and Bush and Mosteller (1951) have, in describing discrimination learning, included the idea that cues which are uncorrelated with the reinforcement lose control over the responses. Specific predictions may be made for the results of the present experiments. These predictions and their relationship to the outcome of the present experiment will be described following the presentation of the results.

[1] This research was supported in part by Research Grant M-3906A from the United States Public Health Service. Thanks are due Samuel H. Revusky and Charles Land for serving as *E*s for half the *S*s.

EXPERIMENT I

Experiment I served to obtain broad coverage on the variable of number of RRs. Either 0, 10, 30, or 60 RRs were presented before the discrimination problem began.

A second variable resulted from consideration of a problem first noted by Spence (1945). It is possible that *S* could literally ignore the stimuli by "closing his eyes" to them. As a result of the RR *S* might decide that the colors had nothing to do with the problem and might make his responses without adequate receptor stimulation, i.e., without directing his gaze to the colors. The decrement in learning which would subsequently appear would in a sense be an artifact, since the reinforcements, when they came regularly, could not strengthen any S-R relation. In order to evaluate whether such an artifact was occurring half of the *S*s were forced to observe the stimuli. They were instructed to describe each stimulus before making their response. The other half of the *S*s did not receive this special instruction. If RRs retarded performance only in the latter

group, it would suggest, as one possibility, that these *S*s were not always viewing the colors.

Method

Subjects.—The *S*s were 16 Indiana University students in an introductory psychology course taught by the author.

Apparatus.—Four decks of cards were employed, one for each of four problems. Within each deck one of two forms (e.g., a diamond or cross) appeared on each card, and each form was of one of two colors (e.g., purple or tan). The four resulting stimuli were distributed randomly throughout a deck of 120 cards.

The four decks were composed of the following stimuli: a purple or tan diamond or cross; a red or green square or semicircle; a yellow or blue circle or × shape; an orange or black triangle or star.

Design.—The observing-instruction variable and 2 different *E*s were organized factorially to produce four different conditions. Four *S*s were assigned to each of the four conditions. Each *S* received four discrimination problems preceded by either 0, 10, 30, or 60 RR trials. The 4 *S*s within any one condition received the four RR amounts in four different sequences counterbalanced to produce a latin square. A different latin square was employed for each condition.

Procedure.—Each *S* was run alone by 1 of the 2 *E*s. The *E* presented *S* with one of the four decks face down, and showed him four different sample cards which were the types making up the whole deck. The *S* was instructed to turn the cards in the deck one at a time and to "sort" the cards into two groups by making one of two responses to each card. The two responses were given as VEC and NONVEC. The *S* was further told that after each response *E* would tell him whether he was right or wrong and that his task was to solve the problem in the sense of getting as many right as possible. If *S* was in the observing-instruction condition he was also told to state the color and form on the card before making his response.

The experiment then proceeded in two main phases: Phase I, the RR part of the experiment, and Phase II, the discrimination proper. Following each response in Phase I, *E* reinforced according to a previously constructed schedule of "rights" and "wrongs." The two events were in a sequence determined randomly with two restrictions: there were never more than three "rights" (or "wrongs") in succession, and half the outcomes were always "rights."

When Phase I was completed *E*, without comment, or obvious change in behavior, shifted his basis for reinforcement to *S*'s responses to the cards. The particular stimulus-response pairs receiving the word "right" had been designated in advance. For example, the positively reinforced pairs might be purple = VEC, tan = NONVEC. Phase II, therefore, consisted in a two-response discrimination problem with one relevant dimension (color) and one irrelevant dimension (form) and with 100% consistency in the stimulus-response-reinforcement combinations. The problem continued in this way either until *S* reached a criterion of 15 consecutive correct responses or, if criterion performance did not appear, until 50 responses had occurred. Following the attainment of the criterion *E*, again without providing any discriminable cue, presented RRs for the next 10 responses. The purpose of this last detail was to minimize transfer of a set for simple solutions from one problem to the next.

The deck was then collected and replaced by the deck for the second problem. The *S* was instructed that the second problem followed the same procedure as the first but was otherwise a brand new problem having no relation to the preceding. The four sample cards for the new deck were shown and two new responses (ZEB and NONZEB) given. The *S* was again told to get as many right as possible, and, if he was in the observing-instruction condition, was asked to describe the color and form on the card before making his response.

Following Phase II of this second deck, the third and fourth decks followed in sequence with instructions similar to those preceding the second deck.

Results

Trials to criterion and number of errors were taken as measures of performance during the discrimination problem. Both measures yielded similar results and the same conclusions. Only the trials to criterion measure, therefore, will be presented here.

The effects of different numbers of RR trials and the influence of the observing-instruction variable upon these effects are shown in Fig. 1.

Two general results may be seen concerning the RR variable. The 0 condition produces faster learning than the other three conditions and the function appears to be flat between 10 and 60 RRs. This graph suggests that the changes produced by RRs occur during the first 10 RR trials and that beyond this number further RRs do not impede learning. Figure 1 also shows that the observing-instruction variable has little effect, and that what effect it does have is in the direction opposite from that expected by the consideration that the noninstructed *S*s might not look at the stimuli. When *S*s are forced to describe the stimulus before each response, thereby demonstrating that the stimulus has been observed, the performance is slightly worse.

The effect of the observing instruction was evaluated statistically. Total trials to criterion following all four RR amounts were obtained for each *S*, these totals were treated as the cell scores in the 2 × 2 factorial design produced by the *E* and instruction variables, and an analysis of variance was performed on these data. Neither the *E*, the observing-instruction variable, nor the interaction was significant.

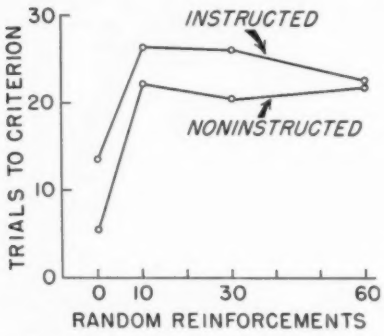

FIG. 1. Mean number of trials to reach criterion following different numbers of random reinforcements, for the group instructed to describe the stimuli and the noninstructed group.

EXPERIMENT II

Experiment I indicated that the value of 10 RRs yielded results which were already at asymptote, that if one wished to show a gradual increase in problem difficulty with an increase in number of RRs, one had to present small numbers of RRs. Experiment I also utilized a latin square design. A problem exists with this design in that the resulting function may describe not only the effects of the variable but may also be influenced by the particular sequences in which the levels of the variable were presented. Accordingly, in Exp. II independent groups were compared in which *S*s received a single problem preceded by either 0, 4, 8, or 12 RRs.

The effects of irrelevant dimensions were also investigated. In Exp. I there was always one relevant (color) and one irrelevant dimension (form) during the discrimination problem. In Exp. II there was again one relevant dimension (color), but different groups received problems with either 0, 1, or 2 irrelevant dimensions.

Method

Subjects.—The *S*s were 144 Indiana University students in an introductory psychology course taught by the author. The *S*s were largely freshmen and were not from the same class as *S*s of Exp. I.

Apparatus.—More than 1,000 3 × 3 in. cards were printed, each containing a single figure. The figure was either a diamond or a cross, which could be either blue or tan and large (1 sq. in.) or small ($\frac{1}{4}$ sq. in.). These cards were divided up into nine decks of 120 cards. All the decks varied with respect to color in exactly the same way: In each set of 20 cards blue and tan each appeared 10 times; neither color appeared on more than three successive cards. The nine decks differed with respect to the number of noncolor

dimensions which were present, as follows: (*a*) 0-decks: There were four different decks which did not vary in any dimension other than color. In one deck, for example, all the stimuli were large diamonds. The other three decks consisted of small diamonds, large crosses, and small crosses, respectively. (*b*) 1-decks: There were four different decks which varied in one dimension independently of the variation in color. In one deck, for example, all the stimuli were large diamonds or crosses. The other three decks consisted of small diamonds or crosses, large or small diamonds, and large or small crosses. (*c*) 2-deck: There was one deck which varied in two dimensions independently of the variation in color. The stimuli could be large or small crosses or diamonds.

All the irrelevant dimensions had the same sequential characteristics in all decks: Each figure (or size) appeared 10 times in 20 cards; no more than 3 cards in succession had the same figure (or size).

Design.—The variables of number of RR trials (0, 4, 8, or 12) and number of irrelevant dimensions (0, 1, or 2) were arranged factorially to produce 12 different conditions. Twelve different *S*s were used in each condition. Within each of the 0 and 1 irrelevant-dimension conditions 3 of the *S*s each received one of the four available decks.

Procedure.—The *E* presented *S* with the appropriate deck face down and showed him the different cards making up the deck (two cards in the 0-decks, four cards in the 1-decks, and eight cards in the 2-deck). The *S* received the same instructions as *S*s in Exp. I. No special observing instructions were given to any *S*, however.

Following each response in Phase I *E* reinforced according to a previously constructed schedule of "rights" and "wrongs." The two events were in a sequence randomly determined with the same restrictions as in Exp. I. An additional restriction was that the last response in Phase I was always called "wrong."

The shift to the Phase II procedure was without detectable change in *E*'s behavior. The relevant dimension was always color. Phase II ended either when *S* reached a criterion of 15 consecutive correct responses or when 100 responses had been made, whichever was less.

Results

The overall mean trials to criterion are shown in Fig. 2. The preliminary

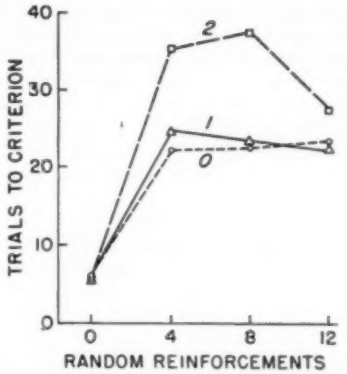

FIG. 2. Mean number of trials required to reach criterion following different numbers of random reinforcements. (The parameter is the number of irrelevant dimensions present during discrimination learning.)

RR trials produce the same two effects that appeared in Fig 1: The 0 condition produces faster learning than the other three conditions, and these latter conditions do not appear to differ among themselves. These results are reflected in the statistical tests. Analysis of variance of all 12 cells shows that the number of RRs is significant. Analysis of the 9 cells of the nonzero conditions shows that the RR variable is not significant.

Figure 2 also suggests that the number of irrelevant dimensions affects 0 RRs differently from the other amounts of RR. The 0, 1, and 2 irrelevant dimensions do not appear to produce differences under 0 RRs, whereas 2 irrelevant dimensions appear to produce greater difficulty for the *S*s receiving 4, 8, or 12 RRs. It should be added, however, that these results are not corroborated by the statistical tests: The interaction between RR and irrelevant dimension is not significant.

DISCUSSION

Figures 1 and 2 show that 4 or more RRs produce a considerable mean per-

formance decrement. Also, if procedural differences between the two experiments may be ignored, when one irrelevant dimension is present this decrement appears to be constant for the range of 4 to 60 RRs.

A plausible interpretation of the decrement is that the RRs have rapidly caused *S* to ignore color as a simple basis for solution. This interpretation is supported by verbal reports of *S*s. In less formal experiments than those reported above, *E* has interrupted *S*s who, after a few RRs and a large number of discrimination trials, had not yet reached criterion. The *E* would then ask *S* what he thought was involved. The *S* invariably indicated that he was attending to a complex sequence (e.g., "say VEG to blue twice then say NONVEG; do just the opposite for tan," or, "it is going VEG, VEG, NONVEG, VEG, NONVEG, but it sometimes changes"). When *E* pointed out that all blue cards were NONVEG, all tan cards were VEG, *S* frequently remarked that he had tried that at the beginning but that it hadn't worked.

The *S*, then, would appear to be ignoring color as a direct basis for solution. The process by which an *S* comes to ignore cues has been treated, in stimulus sampling theory, as a process of cue neutralization. Serious problems occur, however, when this process, as it has been embodied in recent theorizing, is applied to the present results.

Restle (1955) characterized a neutralized cue as "one which the subject does not consider in deciding upon his choice response" (p. 12). Cues, according to his model, gradually become neutralized when they are irrelevant, i.e., when the reinforcements are uncorrelated with the presence or absence of particular responses to those cues. He defined the process of cue neutralization (or, in his earlier terminology, cue adaptation) with this sentence: "If $a(K, n)$ is the probability that irrelevant cue K has been adapted at the beginning of the nth trial, then:

$$a(K, n+1) = a(K, n) + \theta[1 - a(K, n)]$$

is the probability that it will be adapted by the beginning of the next trial" (Restle, 1955, p. 12).

The equation shows that neutralization depends upon the parameter, θ, which is defined as the proportion of relevant cues (cues which, according to Restle, can be used by the subject to predict where or how reward is to be obtained). Its role in the equation is such that cues do not become neutralized when $\theta = 0$, a condition which holds when there are no relevant cues. During the RR phase, in the experiments reported here, color is an irrelevant cue (reinforcements are random with respect to responses made to the colors) and $\theta = 0$ (there are no relevant cues because of the random character of the reinforcement sequence). Because of the latter feature the model predicts that color will not become neutralized, that *S*, after a large number of RRs, will be no more likely to ignore color than after 0 RRs. Figures 1 and 2, however, clearly reveal a neutralization effect.

The same idea is also suggested by Bush and Mosteller (1951). The operator, *D*, which may be interpreted as the mathematical analogue of stimulus neutralization, does not exist when the discriminanda are indiscriminable, i.e., during a random reinforcement series. This operator is assumed only when relevant cues are present. The present results do not confirm this formulation. They suggest, rather, that the models will have to be revised to permit neutralization to be independent of cue relevance.

Performance during Phase II poses a second problem for these theories, for they assume that the process of cue neutralization is irreversible. Once cues are neutralized they are permanently ignored, i.e., they are no longer available for conditioning. Experiment II implies, however, that the color cues are not permanently neutralized. Of the 108 *S*s receiving four or more RRs, only 1 failed to solve the discrimination problem within 100 trials.

In short, the present results suggest that cues are neutralized independently

of the existence of relevant cues, and that the sampling probability of a neutralized cue, while it may become reduced, does not necessarily become zero.

Summary

The effects of random reinforcements (RRs) upon discrimination learning were investigated in two experiments. Two colors were presented in a random sequence and *S* made one of two responses each time a colored stimulus appeared. The *E* said "right" or "wrong" after each response, first in a prearranged sequence (the RR phase), then contingent upon the occurrence of particular stimulus-response pairs (the discrimination phase). In Exp. I the effects of 0, 10, 30, and 60 RRs upon discrimination learning were compared. Experiment II compared the effects of 0, 4, 8, and 12 RRs. Random reinforcements always produced a decrement in performance on the discrimination problem. The decrement appeared to be constant from 4 to 60 RRs. In contrast to the cue-neutralization process described by mathematical models of discrimination learning, the results suggest that cue neutralization does not depend upon the existence of relevant cues, and that the sampling probability of a neutralized cue does not necessarily become zero.

References

Bourne, L. E., & Restle, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66,** 278–296.

Bush, R. R., & Mosteller, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58,** 413–423.

Restle, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62,** 11–19.

Spence, K. W. An experimental test of the continuity and noncontinuity theories of discrimination learning. *J. exp. Psychol.*, 1945, **35,** 253–266.

(Received February 6, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 444-451

REVERSAL AND NONREVERSAL SHIFTS IN DISCRIMINATION LEARNING IN RETARDATES[1]

BETTY HOUSE AND DAVID ZEAMAN

Mansfield State Training School *University of Connecticut*

The second of two discrimination problems may be learned more quickly if the same cues are relevant in both than if different relevant dimensions are presented. This effect has been shown when the possibility of positive transfer of instrumental responses has been ruled out (Lawrence, 1949, 1950) and even, in some cases, when negative transfer of instrumental responses is expected, as in reversal of discriminations (Buss, 1953, 1956; Harrow & Friedman, 1958; Kendler & D'Amato, 1955; Kendler & Kendler, 1959).

The "observing response" model of Wyckoff (1952, 1954) might account for these findings. According to the theory, probability of observing relevant cues increases as a discrimination is learned while probability of observing irrelevant cues decreases. Consequently, relevant cues in a second problem are more likely to be observed if they are the same than if they are different from those in the first problem. The relative frequency of effective over ineffective training trials is thereby increased, and learning is faster when the same cues are relevant in both problems.

The model is thus consistent with the finding that sometimes reversal is learned faster than a new problem with stimuli from a different relevant dimension. A further test is suggested if one assumes, as have Kurtz (1955) and House and Zeaman (1960), that "observing responses" have a broad topography, i.e., may be specified as observation of any value of an entire stimulus dimension such as color or size. It follows that facilitation would also be obtained if the second problem has different relevant cues, but from the *same* relevant dimension.

Some predictions about the expected shapes of learning curves during these transfer tests can also be derived from the theory. Since an observing response must be made to the relevant cue before any learning can occur, an initially low probability will likely produce a number of trials at chance performance preceding the first effective training trial. The result would be the commonly observed ogival discrimination learning curve. Experimental operations designed to decrease or increase initial probability of observing the relevant dimension should have the specific effect of lengthening or shortening, respectively, the initial flat portion of the learning curve.

In the present study, these predictions are tested by training retarded *S*s with a single relevant dimension, then testing under three different shift conditions: (*a*) reversal, (*b*) new problem with cues from the same relevant dimension (intradimensional shift), and (*c*) new problem with cues from a different relevant dimension (extradimensional shift). In addition, two training conditions are employed in which irrelevant cues are either variable or constant. Both training methods have been used by previous investigators (e.g., Kendler & Kendler, 1959; Kendler, Kendler, & Wells, 1960), but no comparison has been made to determine if there are differential effects on performance during the shift tests.

Under the Variable Irrelevant (Variable Irrel.) training condition, two ob-

[1] This investigation was supported by Research Grant M-1099 from the National Institute of Mental Health, Public Health Service. The authors wish to express appreciation to Neil A. Dayton, Superintendent, and to John Cassell and Louis F. Boly, Director of Training and Principal of the Longley School, for their aid and cooperation.

jects differing in color *and* form appear on each trial, e.g. red square vs. green circle. Form might be relevant with the square as positive cue. The square and circle exchange colors randomly from trial to trial. Following such training a group of *Ss* is next split into matched groups to enter the three shift conditions.

Reversal (Rev.) training requires no change in the stimuli. The circle would now be positive instead of the square. Intradimensional Shift (Intra-D Shift) keeps form relevant but shifts to new form cues such as triangle and cross. The old irrelevant colors, red and green, are randomly associated with triangle and cross. Extradimensional Shift (Extra-D Shift) changes the relevant dimension. In the example above, the originally irrelevant red and green cues now become relevant and these randomly exchange two shapes not used in training, such as triangle and cross.

These training and shift conditions provided the experimental design for groups of retardates of low mental age (MA 4–6 yr.). Equivocal results led to a replication with higher MA groups and an added training condition with constant irrelevant stimuli. For the Constant Irrelevant (Constant Irrel.) training condition, the two objects differ in only a single dimension, either color *or* form. An example corresponding to Variable Irrel. above would be a red square vs. red triangle. A reversal was not run following Constant Irrel. training. Intra-D and Extra-D Shift conditions are identical to those after Variable Irrel. training.

Method

Subjects

All *Ss* were institutionalized, mentally retarded boys and girls attending classes at the Longley School of the Mansfield State Training School of Connecticut.[2] Random selections were made within given MA ranges without regard to diagnosis or previous experience. Three training groups may be distinguished as follows:

Variable Irrelevant: MA 4–6.—Thirty-seven *Ss*, 23 boys and 14 girls, with mean MA of 59 mo., mean IQ 41 (range 26 to 62), and mean CA 151 mo. (range 79 to 216), comprised the group.

Variable Irrelevant: MA 6–8.—In this group were 30 *Ss*, 17 boys and 13 girls, with mean MA of 83 mo., mean IQ 56 (range 38 to 69), and mean CA 151 mo. (range 121 to 196).

Constant Irrelevant: MA 6–8.—In this group were 20 *Ss*, 10 boys and 10 girls, with mean MA 85 mo., mean IQ 55 (range 39 to 73), and mean CA 161 mo. (range 104 to 233).

Apparatus

A modification of the Wisconsin General Test Apparatus was used. Components were (*a*) a sliding tray containing two 3-in. food cups centered 12 in. apart and (*b*) a one-way screen separating *E* from *S*. There were 25 different stimulus objects consisting of all combinations of five colors: red, green, yellow, black, and white; and five forms: triangle, square, circle, T, and cross. Forms were cut from ¼-in. Masonite and mounted vertically on 4 × 4 in. gray Masonite bases. Maximum height and width of each stimulus was 2 in.

General Procedure

To bait the food cup, *E* pulled the sliding tray behind the one-way screen, then pushed the tray directly in front of *S* to begin a trial. The *S*'s response was lifting the stimulus object from the concealed cup. Candy rewards were M & Ms, candy corn, and the like. The *E* said "Good" if the first response was correct, and "No" if it was not. Immediate correction of wrong responses was allowed. Position of the correct stimulus, left or right, was varied according to a Gellermann series.

For both training and shift tests, 25 trials per day were given. Criterion of learning was 20/25 correct during a single daily session. For the original problem, training was continued for 125 trials of overlearning beyond criterion. Failure criterion was 250 trials without reaching criterion.

Training

Within each training group, *Ss* were assigned randomly, half to a color-relevant problem and half to a form-relevant problem.

[2] Intelligence test scores were from the Stanford-Binet, 1916 revision, provided by the Psychology Department of the Mansfield State Training School, under the direction of Charles Fonda.

Variable Irrelevant.—From the five colors and five forms available, each *S* was randomly assigned a pair of relevant cues from one dimension, and a pair of irrelevant cues from the other. Thus, a particular *S* might have red and green as relevant cues, square and triangle as irrelevant. To make the irrelevant cues uncorrelated with reward for this *S*, on half the trials the positive cue is associated with one of the two irrelevant cues, and on the remaining trials with the other irrelevant cue. A red triangle vs. green square is alternated over trials with red square vs. green triangle according to a Gellermann series.

Constant Irrelevant.—From either the five colors or the five forms, each *S* was randomly assigned a pair of relevant cues. From the remaining dimension a single cue was randomly chosen. A particular *S* might then have a yellow circle vs. a yellow square for a form relevant problem.

Special Training

For *S*s in the two MA 6–8 training groups, special training was introduced after 250 trials without learning. This procedure was necessary because of a high failure rate. The alternative of training more *S*s and using only learners for the shift tests might produce an artifact. Suppose that *S*s are of two types, those who attend mostly to form and those who attend to color. With a high proportion of failures, there is a possibility that most learners are simply those assigned by chance to their dominant dimensions. Then difficulty upon being shifted to another dimension could be attributed to the introduction of a weaker cue rather than to the effects of training per se. Of the 30 *S*s in the Variable Irrel. group, 12 required special training, as did 10 *S*s of the 20 in the Constant Irrel. group. Since all eventually learned, the possibility of this artifactual explanation is ruled out.

Procedure A.—After 250 failure trials an object differing multidimensionally from the positive cue was selected from a large collection of "junk" objects and sustituted for the negative cue. After criterion, 20/25 correct during a daily session on this simplified problem, regular training procedures were resumed with the original stimuli. Failure criterion was 100 trials for each stage without getting 20/25 correct. Of 22 *S*s receiving this training, 9 continued to fail and were given the following training.

Procedure B.—Without mentioning the color or form of the cues, *E* showed the positive stimulus objects to *S* saying, "These blocks are always right. The candy is always under these blocks." The two incorrect blocks were then presented by *E* with the words, "These blocks are always wrong. The candy is never under these blocks." The appropriate modification was made for the Constant Irrel. group since only one positive and one negative stimulus were used. After these verbal instructions, regular training was begun with no further help from *E*. All *S*s then met criterion of 20/25 correct and received 125 trials of overlearning.

Shift Conditions

The general procedure for the shift tests was the same as for training. The color-relevant and form-relevant training groups were evenly divided among the three shift conditions. The three shift groups were approximately equated for speed of original learning. In particular, *S*s receiving special training were equally represented under each shift condition.

Reversal.—Only *S*s from the Variable Irrel. training condition were given the Reversal test. The same stimuli were used for both training and test. Only the reward values were reversed.

Intradimensional Shift.—The *S*s having Variable Irrel. training were assigned randomly a new pair of relevant cues combined with the same pair of irrelevant cues. The *S*s having Constant Irrel. training were assigned two new relevant cues to be combined with two new irrelevant cues replacing the single cue present during training. Both training groups had the irrelevant cue varied during the shift test.

Extradimensional Shift.—For *S*s having Variable Irrel. training, the pair of cues which were irrelevant during training were made relevant. Two new cues were randomly assigned from the previously relevant dimension to become irrelevant during the shift test. For *S*s having Constant Irrel. training, two new cues from the previously irrelevant dimension were assigned as relevant cues, to be combined with two new cues from the previously relevant dimension.

Results

Variable Irrelevant: MA 4–6.—Only 16 of the 37 *S*s in this group learned the original problem within 250 trials. In spite of small *N*, differences among the shift conditions were significant.

Geometric mean errors to criterion were 6.3, 6.6, and 72.5 for Rev., Intra-D Shift, and Extra-D Shift, respectively, ($F = 5.11$; $P < .05$). The difference between Rev. and Intra-D Shift performance was not significant ($t < 1$), but the Extra-D Shift group was inferior to both of the other groups ($t = 2.6, 2.7$; $P < .05$). These analyses were carried out using log transforms of the error data.

Shift effects for the MA 4–6 group, though significant, were not conclusive because color and form were not equally represented among the shift conditions. Only 4 *Ss* (out of 19) learned an original color problem, while 12 (out of 18) learned a form problem. As a result, both the Rev. and Intra-D Shift groups were mostly *Ss* remaining on a form problem for the shift, while those in the Extra-D Shift group were shifting mostly from form to the more difficult color problem. The ambiguity of these results led to a replication of the experiment with brighter *Ss* and the special training methods required to ensure an unbiased sample on the shift tests.

Variable Irrelevant: MA 6–8.—The ordering of conditions for the MA 4–6 group was reproduced with the MA 6–8 group, with a larger N and an equal number of *Ss* trained with color and form relevant under each shift condition. Geometric mean errors were 6.2, 4.4, and 23.2 for Rev., Intra-D Shift, and Extra-D Shift groups, respectively. Table 1 shows these results broken down by cues. A two-way analysis of variance (Shift × Cue relevant during shift) of log errors to criterion showed the three shift groups differed significantly ($F = 4.79$; $P < .05$). Neither cue during shift nor interaction of cue with shift produced significant effects. Again, the Extra-D Shift group performed more poorly than the Rev. group ($t = 2.38$; $P < .025$) and the Intra-D Shift group ($t = 3.00$; $P < .01$), but the Rev. and Intra-D Shift groups were not different ($t < 1$). The three groups did not differ significantly with respect to speed of original learning ($F = 1.02$; $P > .20$). Geometric mean errors to criterion or (for *Ss* requiring special training) during 250 trials of failure were 30, 38, and 16, respectively, for the Rev., Intra-D Shift, and Extra-D Shift groups. The 12 *Ss* requiring special training were evenly divided among the three shift conditions, and their performance scores showed the same effects as the total group.

TABLE 1

GEOMETRIC MEAN ERRORS TO CRITERION FOR 10 GROUPS OF *Ss* DURING 10 DIFFERENT TRANSFER TESTS

Irrelevant Cue during Training	Conditions during Transfer Tests					
	Reversal		Intra-D Shift		Extra-D Shift	
	Form (F)	Color (C)	F→F	C→C	F→C	C→F
Variable	9.6	4.0	3.1	6.3	20.4	26.3
Constant	—	—	1.8	11.5	44.7	2.0

Constant Irrelevant, MA 6–8.—The difference between the Extra-D Shift and Intra-D Shift groups after constant irrelevant training was not found significant ($F = 1.17$; $P > .25$). The geometric means of the two shift conditions were 4.6 and 9.3 errors to criterion for Intra-D Shift and Extra-D Shift, respectively. Table 1 shows the results for color and form separately. A two-way analysis of variance (Shift × Cue relevant during shift) of log errors to criterion showed that color was more difficult than form ($F = 14.12$; $P < .005$), but interaction of Cue × Shift was not

TABLE 2

ANALYSIS OF VARIANCE OF LOG ERRORS TO CRITERION DURING SHIFT FOR INTRA-D SHIFT AND EXTRA-D SHIFT AFTER VARIABLE IRREL. AND CONSTANT IRREL. TRAINING

Source	*df*	*F*
Shift (S)	1	7.53**
Training (T)	1	1.02
Cues (C)	1	9.91**
S × T	1	1.17
S × C	1	<1
T × C	1	6.82*
S × T × C	1	1.69
Error (*MS*)	32	(.3530)

* $P < .05$.
** $P < .01$.

significant ($F < 1$). Differences between the shift groups during original learning were negligible. Geometric mean errors to criterion or 250 failure trials were 61 and 53 for Intra-D Shift and Extra-D Shift, respectively. The 10 *S*s who required special training were evenly divided among the shift conditions and their results were in the same direction as for the total group.

Constant vs. Variable Irrelevant training.—As a further analysis of the effects of training a 2 × 2 × 2 factorial analysis of variance was performed on the scores of the Intra-D Shift and Extra-D Shift groups for both training conditions. The factors were type of shift (Intra-D or Extra-D), type of training (Variable Irrel. or Constant Irrel.), and cue (color or form) relevant during shift. The analysis (Table 2) shows Extra-D Shift significantly more difficult than Intra-D Shift, and color more difficult than form. Training conditions (Variable Irrel. vs. Constant Irrel.) yield no significant main effect but do interact with type of cue relevant during shift. The low value of 2.0 for the C→F Constant condition is largely responsible for the T × C interaction. This effect may have come about because a strong cue (form) suddenly changing from constant value to variable attracts attention (see Discussion).

Learning curves.—The shape of individual learning curves for these *S*s can best be approximated with group data by means of backward learning curves (Hayes, 1953). In constructing backward learning curves, criterion day scores are combined for each *S* regardless of length of previous training. Scores are then averaged for criterion day minus one, criterion day minus two, etc. In Fig. 1 is shown the result of this computation of the three shift groups after Variable Irrel. training. The curves are extended to the left until the median *S* is dropped. Since most of the Rev. and Intra-D Shift groups met criterion the first day (7 out of 10 in both groups) their backward learning curves are for 1 day only. The median *S* in the Extra-D Shift condition took 3 days to meet criterion. It may be observed that the three curves are quite similar during the last five blocks of trials. Over the entire course of learning, performance of the Extra-D Shift group

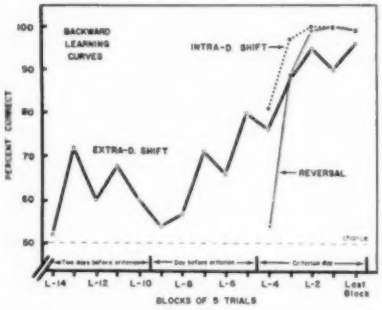

FIG. 1. Backward learning curves for the three shift groups after Variable Irrelevant training. (Curves are extended to the left until the median *S* drops out.)

may be described as ogival while curves for Rev. and Intra-D Shift are negatively accelerated.

As a test of the prediction from Wyckoff's model that if probability of observing relevant cues is decreased, the duration of chance performance will be increased, the first 10 trials of shift learning were analyzed. Figure 2 shows regular learning curves for the first 10 trials for the three shift conditions. It is evident that the Extra-D Shift group remains at chance throughout these 10 trials, while the other two groups show learning effects. A trend analysis showed a significant learning effect ($F = 3.55$; $P < .01$) as well as a significant interaction between Shift and Trials ($F = 4.18$; $P < .01$). Linear components of the shift curves differed ($F = 5.16$; $P < .01$) as did also quadratic components ($F = 6.77$; $P < .01$).

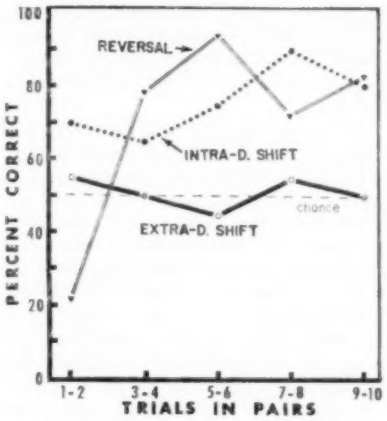

FIG. 2. Learning curves for the first 10 trials for the three shift groups after Variable Irrelevant training.

DISCUSSION

Performance on the shift tests following Variable Irrel. training yields evidence confirming predictions from "observing response" theory. Both the Rev. and Intra-D Shift groups show significant learning effects from the first trial indicating high probability of observing relevant cues. On the other hand, the Extra-D Shift group remained at chance for some trials as would be expected if probability of observing the relevant cues were low.

The superiority of Rev. over Extra-D Shift performance provides grounds for weighing the importance of transfer of a correct observing response. According to observing response theory, *S*s have the starting advantage in reversal of looking at the relevant cues, but the disadvantage of a strong tendency to pick up the wrong one. The *S*s under Extra-D Shift have an initial disadvantage of attending to irrelevant cues, but no corresponding disadvantage of a strong instrumental tendency to pick up the negative stimulus. If overall Rev. performance exceeds Extra-D Shift, therefore, it must be due to positive transfer of observing responses in the Rev. condition, and despite negative transfer of instrumental tendencies. The theory can account for this finding by assigning proper weights to the positive and negative transfer as these combine to determine performance, but the theory can also account for the opposite outcome by changing the weights. Consequently no rigorous test of theory will be provided by comparisons of Rev. and Extra-D Shift until a prior way of fixing parameters is found. The relative importance of observing and instrumental tendencies apparently depends upon the nature of the situation and *S*s, since Rev. has been found sometimes harder than Extra-D Shift and other times easier. Kendler and her associates (Kendler & Kendler, 1959; Kendler et al., 1960) failed to find Rev. faster than Extra-D Shift except for fast learners among 5 and 6 yr. olds, not for slow learners or younger children. For rats, Kelleher (1956) found Extra-D Shift faster than Rev. Reversal has been shown faster than Extra-D Shift for college students with a variety of experimental procedures (Buss, 1953, 1956; Harrow & Friedman, 1958; Kendler &

D'Amato, 1955). Observing response theory can accommodate these varying results but not all theories of discrimination can, even with parameters free for fitting.

The comparison of Variable Irrel. and Constant Irrel. training proved largely indeterminate with the exception of a surprisingly strong performance of the C→F Constant Irrel. group. An interpretation of this finding, related to observing response theory, appeals to the high attention value of novel or suddenly changing aspects of strong stimuli. Form has been shown in this experiment and others to be a stronger cue than color for these *S*s. In the C→F Constant condition form is constant throughout training, then suddenly begins to vary as the relevant dimension during the Extra-D Shift trials. The resulting combination of a relatively strong form cue and novelty apparently makes a highly discriminable dimension. Evidence that novelty or change is a readily discriminable aspect of stimuli for these *S*s has been presented in four previous experiments (House, Orlando, & Zeaman, 1957; House & Zeaman, 1958; Zeaman & House, 1962; Zeaman, House, & Orlando, 1958).

An alternative explanation of the shift effects, highly related to the present theory, has been invoked by Kendler et al. (Kendler & D'Amato, 1955; Kendler & Kendler, 1959; Kendler et al., 1960) to account for the superiority of reversals over extradimensional shift. This theory assumes mediation by implicit responses which add interoceptive cues to the external cues thereby making them more distinct. Kendler and Kendler (1959) have identified these symbolic mediating responses as verbal labels. Observing response theory is similar in that it assumes a mediating response is learned. The difference is that the observing response does not add cues but simply increases the probability that the relevant external cues will be perceived by the organism. The two theories are not incompatible, and it is entirely possible that both effects are present in the data reported here.

However, some of the effects which have been presented are not obviously related to the verbal label theory as presently stated. It is not clear how Intra-D Shift performance would be affected by learning labels which differentiate two specific stimuli. For example, the labels "red" and "green" should have little effect when colors black and yellow are relevant. To handle the Intra-D Shift results, verbal mediation theory might add the assumption that labeling is not a specific tenddency but rather a highly generalized response to an entire dimension of stimuli. Learning to label red and green might then be predicted to generalize to labeling black and yellow or any new colors when presented.

Even with this added assumption, verbal label theory does not share the advantage of observing response theory in predicting the observed shapes of the learning functions.

Summary

Mentally defective children were trained on a discrimination problem with either color relevant and form irrelevant, or the opposite. One group had the irrelevant cue varied randomly during training; the other had a constant irrelevant cue. Following training, *S*s learned either a simple reversal, a new problem with stimuli from the same relevant dimension (intradimensional shift), or a new problem with stimuli from the irrelevant training dimension (extradimensional shift). Extradimensional shift was significantly more difficult than either reversal or intradimensional shift. Reversal and intradimensional shift performance did not differ.

The results were interpreted as consistent with observing response theory, but suggestive of changes for a verbal mediation theory.

References

Buss, A. H. Rigidity as a function of absolute and relational shifts in the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45**, 153–156.

Buss, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.

HARROW, M., & FRIEDMAN, G. B. Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled. *J. exp. Psychol.*, 1958, **55,** 592–598.

HAYES, K. J. The backward curve: A method for the study of learning. *Psychol. Rev.*, 1953, **60,** 269–275.

HOUSE, B. J., ORLANDO, R., & ZEAMAN, D. Role of positive and negative cues in the discrimination learning of mental defectives. *Percept. mot. Skills*, 1957, **7,** 73–79.

HOUSE, B. J., & ZEAMAN, D. Reward and nonreward in the discrimination learning of imbeciles. *J. comp. physiol. Psychol.*, 1958, **51,** 614–618.

HOUSE, B. J., & ZEAMAN, D. Transfer of a discrimination from objects to patterns. *J. exp. Psychol.*, 1960, **59,** 298–302.

KELLEHER, R. T. Discrimination learning as a function of reversal and nonreversal shifts. *J. exp. Psychol.*, 1956, **51,** 379–384.

KENDLER, H. H., & D'AMATO, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49,** 165–174.

KENDLER, T. S., & KENDLER, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58,** 56–60.

KENDLER, T. S., KENDLER, H. H., & WELLS, D. Reversal and nonreversal shifts in nursery school children. *J. comp. physiol. Psychol.*, 1960, **53,** 83–88.

KURTZ, K. H. Discrimination of complex stimuli. *J. exp. Psychol.*, 1955, **50,** 283–292.

LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39,** 770–784.

LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40,** 175–188.

WYCKOFF, L. B., JR. The role of observing responses in discrimination learning. Part I. *Psychol. Rev.*, 1952, **59,** 431–442.

WYCKOFF, L. B., JR. A mathematical model and an electronic model for learning. *Psychol. Rev.*, 1954, **61,** 89–97.

ZEAMAN, D., & HOUSE, B. J. Approach and avoidance in the discrimination learning of retardates. *Child Develpm.*, 1962, in press.

ZEAMAN, D., HOUSE, B. J., & ORLANDO, R. Use of special training conditions in visual discrimination learning with imbeciles. *Amer. J. ment. Defic.*, 1958, **63,** 453–459.

(Received February 9, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 452–457

THE CUE ADDITIVITY PRINCIPLE IN A RESTRICTED SOCIAL INTERACTION SITUATION [1]

JAMES M. RICHARDS, JR.

University of Utah

Recently, increasing attention has been paid to attempts to account for social phenomena in terms of three classic concepts of experimental psychology, namely stimulus, response, and reinforcement (Atkinson & Suppes, 1958, 1960; Burke, 1959; Suppes & Krasne, 1961). The heart of any such attempt is that when two people, A and B, are in a social interaction situation, and A makes some response to the total situation, that response is a stimulus, or cue, for B. When B then responds to this cue originating from A, his response in turn becomes a stimulus, or cue, for A. In other words, each person is a part of the environment of the other person and social cues are in no way qualitatively different from cues originating from other parts of the environment.

Within this general framework there are various ways in which social cues could combine with nonsocial cues, the simplest being through addition. Besides the advantage of simplicity, additive combination of cues offers at least two other major advantages: (*a*) that it makes possible the derivation of *quantitative* predictions through statement in terms of the learning theory developed by Restle (1955, 1957b), and (*b*) that the cue additivity principle has been shown to be a powerful explanatory concept for such diverse phenomena as place-and-response learning (Restle, 1957a), learning sets (Restle, 1958), and concept identification (Bourne & Restle, 1959). There have been, however, no specific tests of the cue additivity principle in social interaction. The present research, therefore, tests two specific hypotheses derived from the application of the general principle of additive combination of cues to social interaction. These are: (*a*) For each person in a two person social interaction situation cues originating from the other person combine additively with other nonsocial cues in the situation. (*b*) If a third person enters the social interaction, the situation is changed only in that the total number of cues is increased additively. More specifically, under the conditions of this experiment, where all persons involved are college age females, the number of cues added by each person is the same.

These hypotheses were stated in quantitative form through use of Restle's theory, which describes learning in terms of the single parameter θ. This parameter, θ, basically describes the rate of learning, and is defined as the proportion of the total cues in the situation which are "relevant" cues in the sense that they are "correlated" with reinforcement, or, in other words, that potentially S can use them as discriminatory cues to obtain reinforcement. The basic quantitative relationship tested is that the rate of learning, θ_{s+n}, for a group of Ss for whom both social cues and nonsocial cues are relevant is equal to the sum of the rates of learn-

[1] Based on a dissertation submitted to the faculty of the University of Utah in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ing, $\theta_s + \theta_n$, of a group of *Ss* for whom social cues are relevant and nonsocial cues are irrelevant and of a group of *Ss* for whom nonsocial cues are relevant and social cues are irrelevant.

Method

Subjects

The *Ss* were 130 volunteer female elementary psychology students who received academic credit for participating. These *Ss* were assigned to two conditions; 50 *Ss* were randomly assigned, 10 to each of five parameter-estimating groups, and 80 *Ss* were randomly assigned, 20 to each of four prediction-testing groups.

Apparatus

The apparatus consisted of an upright black panel to which two white geometric figures, a square and a triangle, were permanently attached. The purpose of this apparatus was to define the two responses available to *Ss* in the experimental situation, and to provide the "nonsocial" cues. In addition to these nonsocial cues, there were various "background" cues (irrelevant presumably) originating from the experimental room, lighting conditions, internal states of *Ss*, etc. It is assumed that these background cues are constant for all groups, and do not affect the equations.

Procedure

It was necessary to control the relationship between social cues and nonsocial cues for each *S*. This control was achieved through use of female stooges who responded in accordance with instructions from *E*. The *Ss* were led to believe that the stooges were also *Ss* and at no time, in so far as could be ascertained from observation, was any *S* aware that the other supposed *Ss* in the experiment were, in fact, stooges. The same stooges were used throughout this experiment.

At the outset of the experimental procedure, each *S*, together with either none, one, or two stooges, was seated in front of the upright black panel and *E* was seated behind the panel. The following oral instructions were then given:

> This is a study of group processes. You see in front of you a panel on which are a triangle and a square. On each trial, each of you will say out loud either "triangle" or "square." In other words, you will choose first (*E* pointing to the person who was to choose first under that experimental condition) and say either "triangle" or "square," whichever you want. Then you (again pointing) will choose second and say either "triangle" or "square," whichever you want. Depending on what the two (three) of you choose, together you will be right or wrong, and I will tell you "right" or "wrong." Then you will choose again, and I will tell you "right" or "wrong," and so forth. Of course, what you are trying to do is to get to the point where you will be right all the time. The choices that are right one time may be, but are not necessarily, the choices that were right the time before. Are there any questions?

Any questions *Ss* might have were answered, and then trials were run until *Ss* reached a criterion of 10 successive correct responses.

The five parameter-estimating groups were run first, and θ values were computed from the mean number of errors on Trials 2 to criterion. These θ values were later combined to yield predicted mean number of errors in each of the four prediction-testing groups. These predicted means were then compared to the observed means of these four groups.

Parameter-Estimating Conditions

In four of the five parameter-estimating groups, only one stooge was present, and in the fifth no stooge was present. The five procedures are as follows:

Group 1.—The *S* chose first on each trial, and the stooge chose second. Nonsocial cues were relevant for *S*, and therefore the symbol for the learning rate parameter in this group is θ_n. The correct response for *Ss* in this group was to choose a designated one of the two geometric figures and to ignore the responses of the stooge. For a randomly designated half of these *Ss*, the figure chosen on Trial 1 was correct throughout, and for the other half the figure chosen was incorrect throughout. In this condition, the stooge chose in a simple alternation pattern, and, on Trial 1, chose the same figure chosen by *S* for half of the *Ss*, and the opposite figure for the other half of the *Ss*. This simple alternation pattern was chosen because it insured that social cues were uncorrelated with rein-

forcement, and therefore irrelevant, while at the same time insuring that the stooge's response on a given trial could be predicted perfectly from knowledge of her immediately preceding response.

Group 2.—The *S*s chose first on each trial and the stooge second, but with social cues relevant and nonsocial cues irrelevant. That is, reinforcement was given whenever *S* chose the same figure as the stooge regardless of what that figure was. The symbol for the learning rate parameter, therefore, is θ_s.

It was originally intended to have the stooge choose in a simple alternation pattern in this group too. Pretests involving 7 *S*s indicated, however, that this was not a satisfactory procedure, since *S*s tended, after a few trials, to fixate on a given figure and therefore never learn. Inquiry revealed that these *S*s were aware that reinforcement was given whenever they and the stooge chose the same figure, but had the hypothesis that the stooge was supposed to follow their choice. The following procedure was therefore used for this group. The stooge would choose in a simple alternation pattern for a block of 6 trials. After this block, if *S* chose the same figure for two successive trials, so would the stooge, thus providing information on the second of these trials disconfirming the hypothesis previously described. The stooge would then return to the simple alternation procedure for another block of 6 trials, etc. This pattern was repeated until *S*s were correct on 10 successive trials. On Trial 1, the stooge chose the same figure chosen by *S* for half of the *S*s, and the opposite figure for the other half of the *S*s.

Group 3.—The stooge chose first on each trial and the *S* chose second. The stooge chose in a 1212122121212112 pattern with the figure designated as 1 for *S* being randomly determined, but with the restriction that each figure be 1 for half of the *S*s. This pattern of responding was chosen for the stooge because it included all possible sequences of four responses for which the responses are divided 50-50 on each even numbered trial, thus making it possible to arrange reinforcement in such a way that social cues are irrelevant.

[1]In this group, reinforcement was given when *S* ignored the choice of the stooge and chose a randomly designated one of the two geometric figures. That is, nonsocial cues were relevant.

Group 4.—The procedure was identical with that followed in Group 3, except that social cues were relevant and nonsocial cues irrelevant.

Group 5.—No stooge was present, and *S*s were reinforced for choosing a randomly designated one of the two geometric figures, with the restriction that each of the two figures was correct for half of the *S*s in this group. It was necessary to run this group in order to derive predictions for the three-person situation.

Prediction-Testing Conditions

Group 6.—One stooge was used, and *S* chose first on each trial. For a randomly designated half of the *S*s, the figure chosen on Trial 1 was correct, and for the other half incorrect. The stooge always chose the correct figure, so that both social cues and nonsocial cues were relevant. This made it possible to test predictions derived from the theoretical relationship $\theta_{s+n}=\theta_s+\theta_n$ where θ_n and θ_s were derived from the parameter-estimating groups in which *S* chose first.

Group 7.—The procedure was identical with that for Group 6 except that a second stooge was added to the situation, who chose after both *S* and the relevant first stooge had chosen. The responses of the second stooge were irrelevant, and were made in the 1212122121212112 pattern. Predictions for this group were derived from the assumption that the second stooge added the same number of cues to the two-person situation as the first stooge added to the situation in which no stooge was present.

Group 8.—One stooge was present who chose first and always chose a randomly designated one of the two geometric figures, with the restriction that each figure was chosen with half of the *S*s in this group. The *S*s were reinforced for choosing the same figure as the stooge chose, so that both social and nonsocial cues were relevant. Predictions for this group were based on the parameter-estimating groups in which the stooge chose first.

Group 9.—The procedure was identical to that for Group 8 except that a second stooge was added who chose before both the relevant first stooge and *S*. This second stooge chose in the 1212122121212112 pattern, and provided irrelevant cues. Again predictions were derived from the assumption that both stooges added the same number of cues.

In all of the above groups it proved necessary, because of scheduling considerations, to eliminate and replace all *S*s who did not learn to criterion within 150 trials. One *S*

was replaced, therefore, in each of four groups: Group 1, Group 2, Group 5, and Group 7. Two *S*s were also eliminated because of errors by *E* in giving reinforcement.

Results

The basic data for all groups in this experiment are the mean numbers of errors made from Trial 2 to the 10 criterion trials. Trial 1 is eliminated because on that trial *S*s had not received any information about which cues were correlated with reinforcement.

A value for θ corresponding to the mean for each of the five parameter-estimating groups was determined from a computing chart (Bourne & Restle, 1959) which presents θ as a function of mean number of errors to criterion. These means and their associated θ values are presented in Table 1.

The discrepancies between predicted and observed means for the prediction-testing groups were tested through use of the t test for significance of deviation from a point prediction. Bourne and Restle (1959) point out that this is the only convenient test currently available for such comparisons, although it is known to be, in some respects, inappropriate. Fortunately for the purposes of this study, the lack of appropriateness results in a conservative error, since a true theory would be rejected too often. Results of these tests are presented in Table 2.

TABLE 1

Mean Number of Errors on Trials 2 to Criterion and Associated θ Values for the Five Parameter-Estimating Groups

Group	Cues	Stooge	Mean	θ
1	Nonsocial, Relevant	None	2.80	.310
2	Nonsocial, Relevant	Chooses 1st	6.10	.171
3	Nonsocial, Relevant	*S* chooses 1st	11.60	.110
4	Social, Relevant	Chooses 1st	3.10	.281
5	Social, Relevant	*S* chooses 1st	5.40	.187

Although none of these differences was significant at the .05 level, there appears to be a reasonably large discrepancy between the predicted and observed values in Group 7. It appeared desirable, therefore, to try to determine whether this discrepancy was due to 1 or 2 *S*s or a definite tendency for the group as a whole to make more errors then predicted. As a partial check, therefore, a sign test was computed in which *S*s in this group making more errors than the predicted mean were given a plus and *S*s making fewer errors than the predicted mean were given a minus. In this group, 12 *S*s made more errors than the predicted mean and 8 *S*s made fewer, a result that is

TABLE 2

Comparison of Predicted and Observed Mean Number of Errors on Trials 2 to Criterion for the Four Prediction-Testing Groups

Group	Condition	Predicted θ	Predicted Mean	Observed Mean	Observed SE_m	t
6	2-Person, Stooge 1st	.452	1.71	1.45	.2350	1.11
7	3-Person, Stooges 1st	.312	2.79	4.35	.7721	2.02
8	2-Person, *S* 1st	.297	3.00	3.30	.5264	.57
9	3-Person, *S* 1st	.181	5.05	5.15	1.1434	.09

Note.—For all t's, $df = 19$; none is significant at the .05 level.

TABLE 3

COMPARISON OF PREDICTED AND OBSERVED MEAN NUMBERS OF ERRORS ON TRIALS 2 TO CRITERION IN THE FOUR PREDICTION-TESTING GROUPS WHERE EACH OBSERVED MEAN IS COMPARED TO AN INAPPROPRIATE PREDICTED MEAN

Group From Which Predicted Mean is Taken	Group From Which Observed Mean is Taken	Predicted Mean	Observed Mean	Observed SE_m	t
3-Person, Stooges 1st	2-Person, Stooge 1st	2.79	1.45	.2350	5.70**
3-Person, *S* 1st	2-Person, *S* 1st	5.05	3.30	.5264	3.32*
2-Person, Stooge 1st	3-Person, Stooges 1st	1.71	4.35	.7721	3.42*
2-Person, *S* 1st	3-Person, *S* 1st	3.00	5.15	1.1434	1.88

* $P < .01$.
** $P < .001$.

obviously far short of significance at the .05 level.

Since none of the differences presented in Table 2 was significant at the .05 level, it also was felt that a check should be computed to determine whether, under the conditions of this experiment, this statistical procedure does typically lead to a rejection of the null hypothesis when *inappropriate* predictions are assessed. Accordingly, t tests were computed in which observed means for the three-person groups were tested against the predicted means for the two-person groups, and vice versa. Results are presented in Table 3.

Finally, in order to determine if the four prediction-testing groups differed reliably among themselves, a fixed constants model analysis of variance (McNemar, 1955) was computed with the rows representing order of choosing (i.e., whether *S* or the stooge chose first) and the columns representing the number of stooges present. In this analysis of variance, the dependent variable was the number of errors on Trials 2 to criterion for each *S*. Results are presented in Table 4.

TABLE 4

ANALYSIS OF VARIANCE OF NUMBER OF ERRORS ON TRIALS 2 TO CRITERION FOR THE FOUR PREDICTION-TESTING GROUPS

Source	*df*	*MS*	*F*
Order of choosing	1	35.11	3.53
Number of stooges	1	112.81	11.33*
Interaction	1	5.52	.55
Within	76	9.96	—

* $P < .01$.

DISCUSSION

The major hypotheses in this experiment were that in social interaction, social cues combine additively with other cues in the situation and that under the conditions of this experiment, each person adds the same number of cues as every other person. Through use of Restle's learning theory, these hypotheses were stated in quantitative form, and the key test of their correctness is how accurately this theoretical formulation predicted the mean number of errors on Trials 2 to criterion for the four prediction-testing groups. The results presented in Table 2 indicate that in none of these groups was the observed mean number of errors significantly different from the theoretical prediction, in spite of the fact that the statistical analysis used would reject a true null hypothesis too often. On the other hand, the data presented in Table 3 indicate that when observed means are tested against inappropriate predictions, in two of these groups the differences are significant beyond the .01 level and in a third group, the difference is significant beyond the .001 level. This suggests that

the failure to find significant discrepancies between predicted and observed values in Table 2 cannot be attributed to a lack of precision of the statistical test. Taken together, these results support the two experimental hypotheses and thus the operation of the cue additivity principle in social interaction.

It should be emphasized that in this experiment the "predictions" were, in fact, genuine predictions. The predicted values were computed prior to the collection of any data for groups used to test those predictions. This experiment, therefore, offered the advantage that the design made it possible to disconfirm the theory. The fact that such disconfirmation was not obtained, together with the fact that the predictions were quantitative rather than qualitative, provides additional support for both the operation of the cue additivity principle in social interaction and the power of Restle's theoretical formulation.

The results of the analysis of variance presented in Table 4 indicate that the only variable in the four prediction groups which had a significant effect was the number of stooges present. The fact that this variable had a significant effect tends to support the theory of Restle, since under the conditions of this experiment the second stooge was adding only irrelevant cues and thus should significantly increase the number of errors. It also provides evidence that the results of this study cannot be attributed to a lack of any real differences among the four prediction-testing groups. Order of choosing did not have a statistically significant effect. Since Restle's theory does not lead to predictions about the effect of this variable, this finding is not inconsistent with the theory. The fact that the interaction term in this analysis of variance was not statistically significant is further evidence that cues did combine additively.

Summary

This study tested two hypotheses derived from the application of Restle's mathematical learning theory to social interaction: (*a*) For each person in a two-person social interaction situation, cues originating from the other person combine additively with other nonsocial cues in the situation. (*b*) If a third person is added to the social interaction, the situation is changed only in that the total number of cues is increased additively.

These hypotheses were tested in a group-learning situation, using groups consisting of 1 *S* and either 0, 1, or 2 stooges, with number of errors to criterion as the measure of learning. In no group was the observed mean number of errors significantly different from the predicted mean number of errors derived through use of Restle's theory. The data, therefore, support both hypotheses and provide evidence for the operation of the cue additivity principle in social interaction.

References

Atkinson, R. C., & Suppes, P. An analysis of two-person game situations in terms of statistical learning theory. *J. exp. Psychol.*, 1958, **55,** 369–378.

Atkinson, R. C., & Suppes, P. *Markov learning models for multiperson interactions.* Stanford: Stanford Univer. Press, 1960.

Bourne, L. E., Jr., & Restle, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66,** 278–296.

Burke, C. J. Applications of linear models to person interactions. In W. K. Estes and R. R. Bush (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959.

McNemar, Q. *Psychological statistics.* New York: Wiley, 1955.

Restle, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62,** 11–19.

Restle, F. Discrimination of cues in mazes: A resolution of the "place-vs.-response" question. *Psychol. Rev.*, 1957, **64,** 217–228. (a)

Restle, F. Theory of selective learning with probable reinforcements. *Psychol. Rev.*, 1957, **64,** 182–191. (b)

Restle, F. Toward a quantitative description of learning set data. *Psychol. Rev.*, 1958, **65,** 77–91.

Suppes, P., & Krasne, F. Application of stimulus sampling theory to situations involving social pressure. *Psychol. Rev.*, 1961, **68,** 46–59.

(Received February 10, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 458–463

RESPONSE-MEDIATED GENERALIZATION WITH SIMPLE SKELETAL-MOTOR RESPONSES[1]

STANFORD H. SIMON[2]

Indiana University

Response-mediated generalization (RMG) may be said to occur when a response R_a, which has closely followed a second response R_b in one or more stimulus situations, follows R_b in a new stimulus situation to which only R_b has previously been conditioned. The mediated or generalized response would be R_a and the mediating response would be R_b.

While an eyeblink (e.g., Lumsdaine, 1939; Shipley, 1933, 1935) and verbal behavior (e.g., Birge, 1941; Cofer, Janis, & Rowell, 1943; Foley & Cofer, 1943; Wickens & Briggs, 1951) have been used as mediating responses, the writer knows of no experiments investigating the skeletal-motor response as a mediating response. There are several reasons for doing research in this area. In the first place, since the current learning theories have developed from data acquired predominantly from experimentation on skeletal-motor responses, such investigations would seem to be most readily related to present data on behavior and to present learning theories. Indeed, the prediction of RMG by present S-R theories would seem most applicable to the skeletal-motor response. In the second place, it would seem that some of the variables involved in RMG would be more accessible for intensive investigation if simple skeletal-motor responses were utilized. Such variables as the temporal relationship between the mediating and mediated responses, the duration of these two responses, and their magnitude could be investigated more readily because of the comparative ease with which simple skeletal-motor responses may be recorded.

The present experiment had three purposes. First, following the prediction of S-R theory (e.g., Hull, 1943), an attempt was made to demonstrate RMG with simple skeletal-motor responses serving as both the mediated and mediating responses. Secondly, RMG as a function of the number of pairings of the mediated response with the mediating response was investigated. This followed from the assumption in S-R theory that the chaining of responses takes place by the conditioning of a response to the response-produced stimuli (RPS) of the preceding response. The hypothesized RPSs are thought to function in conditioning in the same way as do external stimuli. Therefore, the strength of conditioning between the mediated response and the RPS of the mediating response would be expected to be a function of the number of times the responses

[1] This report is based upon a thesis submitted in partial fulfillment of the requirements for the PhD degree at Indiana University. The author sincerely appreciates the assistance given him by W. K. Estes.

Microfilm of the complete manuscript, 66 pp., is available for $1.00 from University Microfilms; Ann Arbor, Michigan, Publication Number 5167. The thesis title is "Response-mediated generalization as a function of (*a*) the number of pairings of the mediating response with the mediated response, and (*b*) an external stimulus contingent on the mediating response."

[2] Now with the Psychology Service, Veterans Administration Hospital, Tomah, Wisconsin.

have been paired. Since RMG is dependent upon this conditioning, RMG would be expected to be a function of the number of pairings of the mediated response with the mediating response.

The third problem investigated was the effect of an external stimulus made contingent on the occurrence of the mediating response. The assumption involved here was that the strength of a learned response is a function of the number of stimuli (or strength of the stimuli) to which it is conditioned (Estes, 1950; Hull, 1943, p. 206). If the number of stimuli contingent upon the mediating response were increased, the mediated response could become conditioned to more stimuli; and if these were present when testing for generalization, more generalization would be expected than would otherwise be the case.

METHOD

Apparatus.—The *S* sat at one end of a table and faced the open end of an oblong box mounted on it. A panel board containing three lights located above the box was attached to the other end. Inside the box was a lever operated by *S* with his right hand. A telegraph key was operated by *S* with his left hand. The *S* wore a pair of earphones.

The box was approximately 1×.75×3.25 ft. Mounted perpendicular to the lever at *S*'s end was a handle covered with sponge rubber to reduce tactual cues of movement. (Visual and proprioceptive cues of movement were reduced by having *S* operate the lever in darkness and at arm's length.) The lever was 54 in. long. The handle displacement was 1 in.

The panel board was 25 in. wide and 26 in. high. Two inches from the top and 5 in. from either side were located two $\frac{3}{4}$-in. lights, the left one red and the right one green. A third light, $\frac{1}{2}$ in. in diameter, was located between the other two and was white. The lights were 4.5 ft. from *S*'s end of the table. The table with the apparatus was enclosed in a small lightproof cubicle.

Behind the cubicle and under the lever an Esterline-Angus recorder recorded key press, stimuli, and lever movement. The ratio of handle displacement to recording pen displacement was 1:4.5.

A noiseless switch operated by *E* in *E*'s booth, located behind and to the side of the cubicle, put either tone (S_t), red light (S_r), or green light (S_g) in circuit with a timer which presented the stimuli for .25 sec. at a rate of one every 3 sec. A second switch put the center white light in or out of circuit with the telegraph key. The tone was 800 cps and 15 db.

Both *S*'s and *E*'s cubicles were located in a semisoundproof room. While operating the apparatus *S* sat in darkness.

Subjects.—The *S*s, 120 students in introductory psychology at Indiana University, were under 25 yr. of age and right-handed.

Procedure.—The general experimental design fell into three phases. In Phase 1, *S* was instructed to press the key quickly whenever S_t or S_r occurred (S_g for control group). These were presented in random order for a total of 50 trials.

In Phase 2, *S* practiced operating the lever as well as the key. *All Ss* were instructed to move lever left and back to center for S_r, and right and back to center for S_g. On any trial on which they were to do both, they were to always press the key first and *immediately afterward* move the lever. Only S_r and S_g were presented, randomly and for a total of 30 trials. No instructions regarding lever movement for S_t were given.

Phase 3 was the same as Phase 2 except S_t test trials were included on Trials 7, 10, 12, 16–25. There were 25 trials with S_r and S_g occurring randomly on non-S_t trials, 3 each occurring both before and after Trial 7. Here, RMG is defined as lever movement following key press to S_t in the same direction as lever movement previously associated with key press.

Group 1 followed the procedure described above and was divided into two subgroups: R and G. Subjects from R pressed the key for S_r and S_t; *S*s from G pressed the key for S_g and S_t. Since all *S*s moved the lever left for S_r and right for S_g, it was predicted R and G would differ in direction of lever movement following key press on S_t trials in Phase 3.

To investigate RMG as a function of the number of pairings of the mediated response with the mediating response, two additional groups were added. Group 2 was the same as Group 1 except *S*s did not use the key during Phase 2. Group 3 was the same as Group 1 except during Phase 2. Group 3

held the lever but did not operate it or the key. (No stimuli were presented during Phase 2 for this group; *E* stated he was adjusting the apparatus.) Groups 2 and 3, having less key press lever movement pairings, were predicted to show less RMG than Group 1.

To investigate RMG as a function of an additional stimulus occurring contiguously with the key press, three more groups were used. Group 4 was the same as Group 1 except that it had the white light in circuit with the key during Phases 2 and 3. Group 5 was the same as Group 4 except that the white light did *not* occur on the S_t trials of Phase 3. Group 6 was the same as Group 4 except that the white light occurred only on the S_t trials during Phase 3. Only in Group 4 was there both a chance for the lever movement to be associated with the white light and the white light present on S_t test trials in Phase 3. Therefore, only Group 4 was predicted to show more RMG than Group 1.

Groups 4, 5, and 6, having maximum training as did Group 1, were expected to show more generalization than did Groups 2 and 3.

Each of the six major groups had R and G subgroups with 10 *S*s randomly assigned to each of the 12 subgroups. Phase 1 was the same for all groups. The points of differences are summarized in Table 1.

Scoring.—Measures of direction and extent of lever movement on S_t trials in Phase 3 were taken. These were made in .04-in. units with the aid of a template. The first response of .04 in. or more occurring within $\frac{3}{4}$ sec. after termination of the key press was recorded. (Because of the 1:4.5 ratio between the handle movement and the recorded movement, the criterion for a scorable response corresponds to .009 in. of actual movement of the handle by *S*.) For all *S*s, movements left were scored minus and movements right, plus. On the basis of previous pilot work it was decided beforehand to include only the first six S_t trials for this study.

"Amount of generalization" may be measured in two ways. Method 1 includes both the direction and magnitude of each response. Method 2 involves only the direction of the response. Both procedures involve the algebraic sum of the plus and minus scores over the first six S_t trials, and each has its own advantages and disadvantages. In this study both were used with the restriction that conclusions would be drawn only where differences were obtained at the 5% level of confidence by *both* procedures.

TABLE 1

DIFFERENCES IN CONDITIONS BETWEEN GROUPS

Group	Number of Responses[a]		Number of Pairings[a]	White Light[b]
	Key	Lever		
1	52	36	36	Not present
2	37	36	6	Not present
3	37	6	6	Not present
4	52	36	36	Present
5	52	36	36	Present, except on S_t trials
6	52	36	36	Present only on S_t trials

[a] Total prior to first S_t trial in Phase 3.
[b] Contingent on key press responses in Phases 2 and 3.

RESULTS

The *t* test was made on all comparisons, and these results as well as the scores by both scoring methods are given in Tables 2 through 4. When comparisons were made between the R and G subgroups of each major group, the scoring procedure outlined above was used, i.e., plus scores were assigned to movements to the right and minus scores to movements to the left. For comparisons between major groups, scores of the two subgroups were combined by giving plus scores to movements in the predicted direction (direction of generalization) and minus scores to movements in the opposite direction. This procedure was followed for both scoring methods.

The results will now be considered as they relate to the three problems raised by the experiment.

To test for RMG, Subgroups R and G of each of the six major groups were compared with each other. The means of the R groups were predicted to be in the minus direction (to the left) and those of the G subgroups in the plus direction (to the right).

Table 2 shows that the means of all subgroups, except R of Group 3, bear

TABLE 2

ANALYSIS ON SUBGROUPS R AND G USING FIRST SIX S_t TRIALS

Group	Scoring Method 1					Scoring Method 2				
	Mean		*SD*		*t*	Mean		*SD*		*t*
	R	G	R	G		R	G	R	G	
1	−9.56	11.20	12.46	15.57	3.12**	−1.62	1.90	1.49	2.55	3.59**
2	−11.20	7.30	18.61	13.61	2.40*	−.80	1.20	2.75	2.32	1.66
3	2.60	3.70	29.37	10.30	.11	.85	1.10	2.30	1.64	.27
4	−22.70	15.50	42.13	22.47	2.40*	−.70	2.90	2.61	2.02	3.27**
5	−40.20	20.14	49.90	41.29	2.80*	−1.65	1.78	2.67	2.24	2.96**
6	−8.70	14.34	9.50	9.50	2.51*	−1.80	2.54	1.66	1.80	5.29***

* $.05 > P > .01$.
** $.01 > P > .001$.
*** $P < .001$.

out this prediction with both scoring procedures. Statistical tests with both scoring procedures show a difference between R and G at the 5% level or more for all groups except Groups 2 and 3. Group 2 showed significance at the 5% level only with the first scoring procedure. Neither scoring procedure demonstrated a difference between R and G of Group 3.

In studying RMG as a function of the number of pairings of the mediated response with the mediating response we are concerned with differences between the minimally trained groups, Groups 2 and 3, and the maximally trained groups, Groups 1, 4, 5, and 6. (Groups 4, 5, and 6 also have the light as a variable.)

TABLE 3

GROUP MEANS AND *SD*s USING FIRST SIX S_t TRIALS

Group	Scoring Method 1		Scoring Method 2	
	Mean	*SD*	Mean	*SD*
1	10.38	14.13	1.76	2.09
2	9.25	16.42	1.00	2.55
3	.55	22.23	.13	2.22
4	19.10	33.95	1.80	2.09
5	30.20	46.33	1.72	2.41
6	11.52	19.61	2.17	1.77

Groups 2 and 3 were the only groups of the six major groups that did not show a significant difference between their two subgroups by both scoring methods. The question now remains whether or not the minimally trained groups differ significantly from the maximally trained groups.

The analysis of variance over Groups 1, 2, and 3 was not significant with either Scoring Method 1 ($F = 1.71$, $df = 2/58$, $P > .05$) or Scoring Method 2 ($F = 2.42$, $df = 2/58$, $P > .05$). When the first scoring procedure is used the only group comparison that is significant at the

TABLE 4

RESULTS OF *t* TESTS BETWEEN GROUPS WITH BOTH SCORING METHODS

Group	*t*					
	1	2	3	4	5	6
1		.23	1.67	1.03	1.80	.21
2	1.82		1.37	1.14	1.86	.39
3	2.31*	1.13		1.99	2.52**	1.61
4	.06	1.92	2.39*		.84	.86
5	.05	.89	2.12*	.11		1.62
6	.63	1.64	3.14***	.60	.65	

Note.—Scoring Method 1 is above the diagonal and Scoring Method 2 is below the diagonal.
* $P = .05$.
** $P = .02$.
*** $P = .01$.

5% level is the difference between Groups 3 and 5 (Table 4). The second scoring procedure, which considered only direction of response, showed significant differences between Groups 3–1, 3–4, 3–5, and 3–6 (Table 4). All comparisons were in the predicted direction, i.e., the means of the maximally trained groups were larger than the means of the minimally trained groups, with both scoring procedures.

In studying RMG as a function of an external stimulus occurring contiguously with the mediating response we are concerned only with the groups that received maximum training. The comparisons made were between Groups 4–5, 4–6, 5–6, 1–4, 1–5, 1–6. No significant differences were found in any of these comparisons with either scoring method (Table 4).

Discussion

Of the three hypotheses tested in this experiment, the evidence is clear for only two: the prediction that RMG can be obtained with simple skeletal-motor responses was verified; the prediction that a light made contingent on the occurrence of the mediating response would affect RMG significantly was not borne out. In general, the results seem to favor the third hypothesis, that RMG is a function of the number of pairings of the mediating response with the mediated response. However, since the results are not decisive, no definite conclusion was made.

If there is a real difference between Group 3 and the maximally trained groups, then it is difficult to say why the other minimally trained group, Group 2, did not tend to differ significantly from the maximally trained groups as did Group 3. It may be that while making the lever movement response in Phase 2 of the experiment, *Ss* in Group 2 were also making key-press-like responses with their left hands even though they were not actually operating the telegraph key.

An important problem in this experiment is that of controlling the temporal relationship between the mediating and mediated responses. This may be related to the large variability between *Ss*' scores in this experiment. While the instructions insured the completion of the key press response before the start of the lever movement response, it did not insure the same temporal relationship between these two responses for all *Ss*.

Another problem involves the two scoring methods used in this experiment. The second scoring procedure differed from the first in that it resulted in less within-group variability and provided more significant differences. The first procedure, however, does not sacrifice information about the extent of movement as does the second procedure. If some of the problems involved in this experiment could be resolved with further research, e.g., the variability between *Ss* in their temporal relationships between the mediated and mediating responses, the first scoring method might prove to be the more sensitive in demonstrating the effects of changes in the experimental variables.

Summary

On the basis of current S-R theory, including assumptions concerning the functions of response-produced stimuli, it was predicted that response-mediated generalization (*a*) can occur with simple skeletal-motor responses, (*b*) will be greater for one group of *Ss* than another if the first group has more pairings of the mediated response with the mediating response, and (*c*) will be greater for one group of *Ss* than another if the first group has an additional stimulus occurring contiguously with the mediating response.

These predictions were tested in a simple motor learning situation with human *Ss*. Pressing a telegraph key served as the mediating response and moving a lever served as the mediated response.

The experimental results support the first prediction but not the third; no definite conclusion is drawn regarding the second prediction.

REFERENCES

BIRGE, J. S. Verbal responses in transfer. Unpublished doctoral dissertation, Yale University, 1941.

COFER, C. N., JANIS, M. G., & ROWELL, M. M. Mediated generalization and the interpretation of verbal behavior: III. Experimental study of antonym gradients. *J. exp. Psychol.*, 1943, **32,** 266–269.

ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57,** 94–107.

FOLEY, J. P., JR., & COFER, C. N. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32,** 168–175.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

LUMSDAINE, A. A. Conditioned eyelid responses as mediating generalized conditioned finger reaction. *Psychol. Bull.*, 1939, **36,** 650.

SHIPLEY, W. C. An apparent transfer of conditioning. *J. gen. Psychol.*, 1933, **8,** 382–391.

SHIPLEY, W. C. Indirect conditioning. *J. gen. Psychol.*, 1935, **12,** 337–357.

WICKENS, D. D., & BRIGGS, G. E. Mediated stimulus generalization as a factor in sensory pre-conditioning. *J. exp. Psychol.*, 1951, **42,** 197–201.

(Received February 23, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 464–467

ESCAPE PERFORMANCE AS A FUNCTION OF DELAY OF REINFORCEMENT [1]

HARRY FOWLER AND MILTON A. TRAPOLD

University of Pittsburgh *State University of Iowa*

The present investigation extends our program of studies on factors affecting escape performance (Bower, Fowler, & Trapold, 1959; Trapold & Fowler, 1960) to the delay-of-reinforcement variable. Specifically, the purpose of this study was to assess the effect of different delays of shock termination on the performance of rats running down a charged alley to a goal box that became uncharged at various delays following *S*'s entry into the goal. A wide range of delays in shock termination was investigated in order to provide a general picture of the function relating escape performance to the delay-of-reinforcement variable.

METHOD

Subjects.—The *S*s were 40 male, albino rats, approximately 180 days old at the start of the experiment. All *S*s had served previously as control animals in runway experiments wherein the conditions of food reinforcement imposed upon them were constant over training and uncorrelated with speeds. For a period of 2 weeks immediately prior to the start of the present study, *S*s were housed individually in the laboratory and were maintained on an ad lib. diet of Purina lab checkers and water.

Apparatus.—A straight alley, 68 in. long, 3 in. wide, and 4 in. high, was constructed of two L shaped strips of galvanized sheet metal, and was covered with a transparent Plexiglas top. One of the sheet metal strips formed one side and half of the floor of the apparatus; the other strip formed the other side and other half of the floor. A 1-in. gap separated the two strips at the floor.

The alley comprised three sections, a start box 8 in. long, an alley section 48 in. long, and a goal box 12 in. long. These sections were separated by upward acting guillotine doors. Raising the start-box door operated a microswitch which started a Standard Electric clock and closed a shock circuit to the entire apparatus. Breaking a photobeam 6 in. inside the alley stopped the first clock and started a second which, in turn, was stopped when *S* broke a second photobeam 2 in. inside the goal box. The breaking of the second photobeam also activated a Hunter timer which interrupted the shock circuit to the apparatus only after a specified delay following *S*'s entry into the goal box. The time data from the two clocks were recorded to the nearest .01 sec. and were converted to speed scores, i.e., (100/time). Hereafter, these scores will be referred to as starting and running speeds, respectively.

A single matched impedance, 60 cps, ac shock source provided shock to the apparatus. In this shock system, the galvanized metal strips forming the halves of the apparatus were connected across the output of a power transformer. In addition, a 0.25-megohm resistor was placed in series with *S*. The *S* received shock as long as its feet touched both of the metal strips forming the halves of the apparatus. The narrowness of the alley prevented *S* from running along only one side of the runway.

Procedure.—The *S* was placed in the start box with the start-box door closed and the goal-box door open. When *S* oriented toward the goal, the start-box door was opened and *S* could then run to the goal end of the alley. When *S* entered the goal box, the goal-box door was closed and *S* was detained in the goal box for 30 sec. or longer, as specified in the design section below. At the end of this time, *S* was removed from the goal box and immediately returned to the start box for the beginning of the next trial. In this manner, each *S* received a total of 28 highly massed trials.

Experimental design.—The 40 *S*s were assigned to eight groups of 5 each in such a way as to randomize past experimental history as well as all other extraneous factors. For all groups, intensity of shock in the apparatus

[1] This experiment was conducted at Yale University while H. Fowler held a predoctoral fellowship from the National Institutes of Mental Health, Bethesda, Maryland.

was 240 v.; however, for different groups, this shock was terminated after different delays following *S*'s entry into the goal box and interruption of the final photobeam. Six of the eight groups of *S*s were assigned to one of the following delays in shock termination: 0, 1, 2, 4, 8, or 16 sec.

With termination of shock after the specified delay, *S*s of each of these six groups were detained in the goal box for an additional 30 sec. This procedure provided each of the six groups with the same length of time during which shock was absent at the goal, but confounded total length of time spent in the goal box (and consequently intertrial interval) with the delay variable. To account for this confounding, the two remaining groups of *S*s received different delays in shock termination at the goal, 8 and 16 sec., but with total time spent in the goal compartment held constant at 30 sec. for both groups. This feature of the design permitted possible extraneous effects resulting from differences in total goal-box time and intertrial interval to be assessed by a comparison of the performances of the main and alternate sets of 8- and 16-sec. delay groups.

It should be noted that the concomitant use of both the main and alternate delay conditions was additionally necessitated by the fact that the alternate condition, itself, confounds time during which the shock is absent at the goal with the delay variable. Emphasis was given to the main condition since it was anticipated that the confounding present with this condition would tend to offset rather than facilitate effects of the delay manipulation; the longer intertrial intervals associated with the longer delays should serve to increase speeds (e.g., Teichner, 1952), thereby reducing performance decrements attributable to delay of reinforcement.

Results

Group mean performance curves over the entire course of training are presented for the starting and running speed measures in Fig. 1 and Fig. 2, respectively. To reduce the cluster of curves, the speed data for the alternate 8- and 16-sec. delay groups were not included in these figures. The data for these groups are presented below in relation to appropriate statistical comparisons.

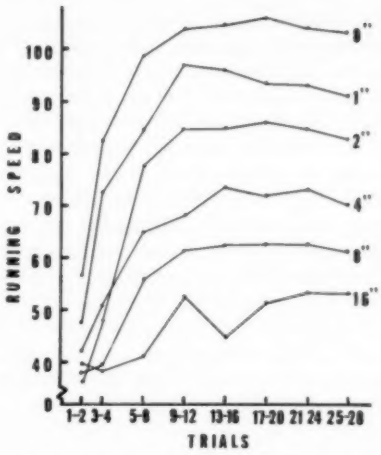

FIG. 1. Running speed (100/time in sec.) as a function of the delay (in sec.) of shock termination.

Running speeds.—Figure 1 shows that differences in mean running speed among the main delay groups emerged early in training, and increased progressively as the various groups approached different levels of asymptotic performance. Because of the generally orderly progression of the running speed curves over the course of training, statistical analysis of group mean differences was limited to the last 12 trials (17–28). The results of an analysis of variance of these data showed that the overall differences among the groups were highly reliable ($F = 6.0$, $df = 5/24$, $P < .001$). In addition, analysis by orthogonal polynomial comparisons of the trends within these data showed that mean running speed over Trials 17–28 was linearly related to the log of the delay values employed ($F = 30.0$, $df = 1/24$, $P < .001$); the residual variance after extraction of this linear-log component was negligible ($F < 1.0$).

As mentioned previously, total time spent in the goal box (and consequently, intertrial interval) are con-

founded with the delay variable for these main groups. However, the speed performances of the alternate 8- and 16-sec. delay groups, for which both total goal-box time and intertrial interval did not covary with delay of shock termination, indicate that these extraneous variables did not contribute significantly to the obtained performance differences. Over the last 12 training trials, mean running speeds for the *alternate* 8- and 16-sec. delay groups were 58.4 and 49.5, respectively. Although these speeds are reliably different from that of the zero or no-delay group ($F=52.1$, $df=1/12$, $P<.001$), they are only slightly and not significantly lower than those for the main 8- and 16-sec. delay groups ($F < 1.0$).

Starting speeds.—Figure 2 shows that the results obtained with the starting speed measure are generally the same as those for running speed. Differences in mean performance begin to emerge early in training and progressively increase until performance becomes asymptotic. Contrary to the running speed results, however, an analysis of variance of group mean starting speeds over the last 12 training trials showed that the overall differences among the delay groups were not significant ($F=2.0$, $df=5/24$, $P>.10$). Nevertheless, trend analysis of these data by orthogonal polynomial comparisons indicated that mean starting speed over these trials, like running speed, was linearly related to the log of the delay values employed ($F=9.7$, $df=1/24$, $P<.005$); again, the residual variance after extraction of this linear-log component was negligible ($F < 1.0$).

The pattern of results obtained for the alternate 8- and 16-sec. delay groups with the starting speed measure was the same as that obtained with running speed. Mean starting speeds for these groups over the last 12 training trials were 165.5 and 148.9, respectively. Again, these speeds are reliably different from that of the no-delay group ($F = 15.1$, $df = 1/12$, $P < .005$), but only slightly and not significantly lower than those for the main 8- and 16-sec. delay groups ($F < 1.0$). Hence, these data also indicate that the obtained differences are not attributable to the covariation of intertrial interval or total goal-box time with delay of shock termination.

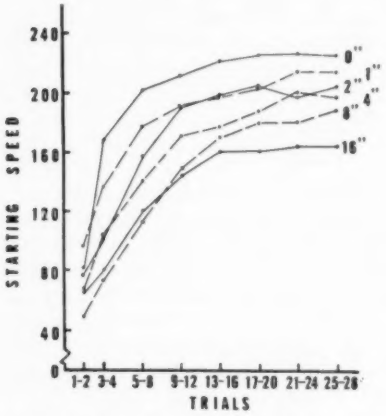

FIG. 2. Starting speed (100/time in sec.) as a function of the delay (in sec.) of shock termination.

DISCUSSION

In total, the results show that for rats running down a charged alley speed is significantly decreased when reinforcement, i.e., shock termination, is delayed following the behavior which is instrumental to it. Further, this decrement in performance is a decreasing, negatively accelerated function of the delay interval, at least over the range of values employed in this study. In terms of the two response measures used, these effects were quite pronounced with running speeds and somewhat less so with starting speeds.

In the context of the behavior theory

developed by Spence (1956) for instrumental appetitive conditioning, the effects of delayed shock termination can be accounted for by the inhibitory intervening variable, I_t, the basis for this inhibition being competing responses which are conditioned during the delay period. In the present study, *S*s were observed to exhibit, following their entry into the goal box and in so far as the length of the delay period permitted, responses such as turning about, cringing, pushing against the goal-box cover, etc. From the competing response theory of I_t, these responses, which become conditioned to the cues at the goal, and hence anticipatory by way of the similarity of the cues in the alley, would tend to interfere with running and thereby reduce speed. In the same vein, the differential effect of the delay manipulation on starting and running speeds can be attributed to a decrease in the similarity of alley and goal-box cues, the greater the distance from the goal. Finally, these inhibitory effects should be more pronounced with longer delays, since the opportunity for the interfering responses to occur and become conditioned will be greater, the greater the length of the delay period.

With the previous studies in this series, additional parallels between appetitive and escape conditioning were demonstrated in the similarity of operation of both the drive and magnitude-of-reinforcement variables. For rats running down a charged alley to a goal box which is either completely or partially uncharged, escape performance is positively related to both shock intensity in the alley (Trapold & Fowler, 1960) and magnitude of shock reduction at the goal, with shifts in the magnitude of goal-shock reduction leading to rapid and appropriate shifts in level of performance (Bower et al., 1959). Taken together, the total findings of this series of studies strongly suggest that escape conditioning can be interpreted in terms of behavior theory formulations now offered primarily for instrumental appetitive conditioning.

Summary

Rats were trained to run from an alley where they received a continuous electric shock of 240 v. to a goal box where shock was terminated at either 0, 1, 2, 4, 8, or 16 sec. following *S*'s entry into the goal compartment. Escape performance was measured in terms of response speeds both at the start and alley sections of the runway. In general, the results showed that speeds were slower the greater the delay of shock termination at the goal. This effect was quite pronounced for speeds in the alley section of the runway, but considerably less so for speeds at the start section. These results are analogous to those obtained in runway situations where delayed appetitive rewards have been employed and, as such, they indicate, along with those findings previously reported, that similar behavioral laws underlie both appetitive and escape conditioning.

References

Bower, G. H., Fowler, H., & Trapold, M. A. Escape learning as a function of amount of shock reduction. *J. exp. Psychol.*, 1959, **58**, 482–484.

Spence, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.

Teichner, W. H. Experimental extinction as a function of the intertrial intervals during conditioning and extinction. *J. exp. Psychol.*, 1952, **44**, 170–178.

Trapold, M. A., & Fowler, H. Instrumental escape performance as a function of the intensity of noxious stimulation. *J. exp. Psychol.*, 1960, **60**, 323–326.

(Received February 24, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 468–471

RESPONSE STRENGTH AS A FUNCTION OF DRIVE LEVEL AND PRE- AND POSTSHIFT INCENTIVE MAGNITUDE [1]

DAVID EHRENFREUND AND PIETRO BADIA

Adelphi College

About ten years ago Hull (1951), largely on the basis of Crespi's (1942, 1944) reports, raised the status and importance of the incentive magnitude variable by removing it as a determiner of H and introducing a separate incentive motivational intervening variable, K, which is designed to handle both differences and shifts in incentive magnitude. The manner in which differences in K combine with other intervening variables particularly D has not yet received enough empirical support to justify, with any large degree of confidence, specifying an exact function. Hull proposed a multiplicative function but, as Spence (1956) has pointed out, this was purely a working hypothesis since there was no evidence.

Although Spence (1956) suggested a linear function (Ramond, 1954), he (Spence, 1960) later suggested that . . . "it is possible that this will have to be changed to some nonlinear function" (p. 96). To date, however, runway studies, at least, have tended to support the additive function (Reynolds & Pavlik, 1960; Weiss, 1960).

The manner in which D combines with shifts in incentive magnitude would also seem to be important theoretically although there have been no reported empirical studies.

The present study involved a 2×2 factorial design involving two levels each of drive and reward. After asymptotic speeds were presumed to be reached, the magnitude of incentive of each group was shifted to that of the other.

[1] Supported by a grant from the National Science Foundation.

METHOD

Subjects.—The *S*s were 20 Sprague-Dawley strain white rats, approximately 100 days old, equally divided among four groups: high drive-low reward, high drive-high reward, low drive-low reward, and low drive-high reward. One *S* in the last group had to be discarded, reducing the N for this group to 4.

Apparatus.—The sides and bottom of the runway apparatus were constructed of cloudy plastic. The top was clear plastic covered with two thicknesses of fine wire screen. The runway was 5 ft. long, and $2\frac{1}{2}$ in. wide, and 4 in. high. It was sandwiched between two large opaque boards which measured $6 \times 2\frac{1}{2}$ ft. The space between these boards was completely enclosed around the outside border. A 1 in. $\times$ 5 ft. slit was made in the bottom board beneath the center of the floor of the maze. With this arrangement, filtered light beams were directed from above the maze to photoelectric cells below the maze. In the present study speed was recorded in the middle 2-ft. section of the maze with a Standard Electric timer to .01 sec. The maze was illuminated by four $7\frac{1}{2}$-w. bulbs along each side. These were cut down by a variac to 40% and the cloudy plastic sides diffused the lights evenly in the maze. Guillotine type cloudy plastic doors, 15 in. from each end of the maze created a start box and goal box. Just in front of this door in the start box there was another horizontally sliding door of clear plastic. Thus walls, floor, and ceiling (except for slits for the doors) were continuous, homogenous, and unbroken from one end of the runway to the other.

Drive and reward.—Drive was operationally defined in terms of a percentage of *S*'s weight under conditions which maintained this weight continuously thus precluding long periods of deprivation. During the experiment all *S*s lived in a weight control apparatus, the details of which are described in a previous paper (Ehrenfreund, 1960). Es-

TABLE 1

INDIVIDUAL GROUP ANALYSIS OF ASYMPTOTIC SPEED SCORES

Reward	Body Weight			
	95%*		85%	
	Mean	*SD*	Mean	*SD*
45 mg.**	10.26	.82	13.92	1.34
260 mg.	13.86	1.27	15.15	1.04

* $P < .005$ for low-drive groups.
** $P < .005$ for low-reward groups.

sentially each *S* is maintained within 1% of a specified weight. Loss of weight in the order of up to 2 gm. results in the automatic delivery of a 97-mg. Noyes tablet every 12 sec. until the specified weight is regained. Water is always available. Under such a schedule *S*s receive small amounts of food aperiodically around the clock. For the purposes of this study the high-drive *S*s were maintained at 85% of ad lib. or starting weight and the low-drive *S*s at 95% of ad lib. weight.

On each trial the low-reward group received a 45-mg. Noyes tablet. The high-reward *S*s received a 260-mg. tablet.

Procedure.—When *S* faced the vertical door of the start box it was raised and 2 sec. later the horizontally sliding clear plastic door was opened. Since under the present drive maintenance schedule *S*s are always at the stated drive level, data could be collected several times a day. All *S*s received 90 acquisition trials under the original drive-reward conditions. On Trial 90, each *S* in a high-reward group (260 mg.) was shifted to a low reward (45 mg.), and vice versa for 25 post-shift trials. Trials were run in blocks of 5 with 15 to 20 sec. between trials and from ½ to 1 hr. between 5-trial blocks. Between blocks of trials, *S*s were returned to the weight control apparatus. All of the 115 trials were run over a 4-day period; 20 on Day 1, 30 on Days 2 and 3, and 40 on Day 4. The median of 5-trial blocks for each *S* was converted to its reciprocal. The means of the scores thus converted form the bases for the tables, graphs, and statistical analyses.

RESULTS AND DISCUSSION

Prereward shift considerations.—The last two blocks of trials before the shift in reward magnitude were assumed to reflect asymptotic values.

An analysis of variance of these values revealed that the main effects of Drive and Reward were both significant beyond the .001 level, the *F* values being 19.85 and 18.90, respectively. The *F* value for interaction was 4.52 which is significant at the .05 level. Table 1 supplies the analysis for the individual group means. Note that the difference between the two low-drive groups (or low reward) is greater than that between the high-drive groups (or high reward). A simple plot of these asymptotic means reveals that the departure from linearity is converging as demanded by exponential addition rather than diverging, as demanded by a multiplicative function.

In order to facilitate comparison of the acquisition curves, the four groups are paired in different combinations in Fig. 1 and 2. The delayed separation of the curves for the two high-drive groups, in contrast to those of the two low-drive groups, lends further support to a negatively accelerated function. Under conditions of either high drive or high reward, the differential increments of combined *D* and *K* would be small.

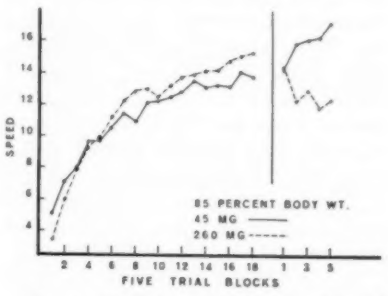

FIG. 1. High-drive groups rewarded with either 45 mg. or 260 mg. (To the right of the vertical line are shown the effects of changing the reward of each group to that of the other group.)

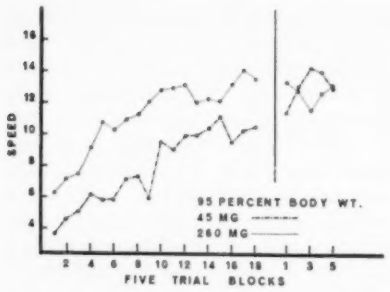

FIG. 2. Low-drive groups rewarded with either 45 mg. or 260 mg. (To the right of the vertical line are shown the effects of changing the reward of each group to that of the other group.)

On the assumption that habit strength multiplies with this combination, $E = H \times (D + K)$, the separation of the groups would not be manifest until H had developed some appreciable magnitude. These data are not in agreement with two recent studies (Reynolds & Pavlick, 1960; Weiss, 1960) which failed to find an interaction between drive and reward and therefore supported a linear or simple additive function.[2] Since both of these studies employed hours of deprivation as contrasted with the weight control method of the present study, the discrepancy might be attributed to the different operations defining drive. We feel, however, that there is perhaps a more significant, albeit more subtle, difference which relates to response measurement. Both the Weiss (1960), and Reynolds and Pavlick (1960) speed scores included goal-box speeds; whereas the speed scores of the present study derive from the middle of the alley, uninfluenced by either starting or goal-box speeds. The advantages of delineating at least these three response speeds (i.e., starting, running, goal) has recently been demonstrated by Goodrich (1959) and Wagner (1961) whose data are particularly convincing in reference to goal-box versus running speeds.

The effect of shift in incentive magnitude.—The immediate effect of the shift in magnitude was an increase in general excitement for practically all *S*s, as revealed in increased difficulty of handling, resistance to being placed in the start box, and, in some cases, a delay in the consummatory response even with those *S*s shifted upward. These behaviors were also reported by Crespi (1944) who, however, only mentioned them in connection with *S*s shifted downward. Not only did we observe them for *S*s shifted upward (although to a lesser extent) but more significantly, in terms of the analysis of the present data; they were clearly more obvious in the high-drive *S*s.

The right sides of Fig. 1 and 2 show the effect of shifting reward from 260 mg. to 45 or from 45 mg. to 260. Our main interest is centered on the fact that both the elation and depression effects obtain under high drive ($P < .05$), but neither obtains under low drive. Assuming this to be a genuine phenomenon, the problem remains as to how to account for it within the theory. The evidence to date is much too sparse to justify a detailed theoretical analysis. Nevertheless, existing theoretical concepts elaborated in other experimental contexts, coupled with qualitative observations of *S*s' behavior mentioned above, suggest the importance of the role of excitement, or the fractional anticipatory emotional response r_e as developed by Spence (1956, 1960) or r_f as developed by Amsel (1958).

As the data now stand, no new assumptions other than those already developed concerning the role of r_e are required. Occurring originally in the goal box, r_e

[2] It might be pointed out, however, that the data of the present study are exactly in line with the type of additive hypothesis that Spence (1953, p. 20) proposed in the Kentucky Symposium.

would generalize to the highly similar runway. Also r_e tends to elicit through its response produced cues ($r_e \rightarrow s_e$), overt responses, some of which may be incompatible with the running response. Thus the occurrence of r_e in combination with the existing drive level will increase or decrease response strength to the extent that experimental manipulations tend to facilitate or preclude the development of such competing responses of varying strengths. In the present situation it is hypothesized that such incompatible responses are much more likely to occur under conditions of reduction in incentive magnitude. This will be recognized as essentially the same theoretical analysis made by Spence (1960) for extinction phenomena which are here regarded as the limiting case of incentive magnitude reduction. Accordingly, the postshift response decrements and increments are seen as functions of drive level and the difference and direction between pre- and postshift reward magnitudes. In these terms, the so called elation and depression effects would become merely points on a curve and thus disappear as independent phenomena.

Summary

In an attempt to ascertain (*a*) the function describing how D and K combine and (*b*) the effect of incentive magnitude shifts as a function of D, speed measures were obtained on four groups of rats involving two drive levels and two reward levels. After 90 trials on the original reward level, the reward of each group was shifted to that of the other for 25 additional trials. Statistical analyses revealed a significant interaction ($P = .05$) between D and K, indicative, not of a multiplicative function but of a negatively accelerated one such as exponential addition. Both main effects were highly significant but t tests were significant only between the two low-drive groups and between the two low-reward groups. After the shift in incentive magnitude, changes in running speed occurred in the expected direction but the differences were much greater under high drive than under low drive. It was hypothesized that such differences (i.e., Crespi effect) were an increasing function of D and the extent of the shift in incentive magnitude. The shift data were interpreted in terms of the hypothetical anticipatory emotional response mechanism ($r_e \rightarrow s_e$).

References

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

Crespi, L. P. Amount of reinforcement and level of performance. *Psychol. Rev.*, 1944, **51**, 341–357.

Ehrenfreund, D. The motivational effect of a continuous weight loss schedule. *Psychol. Rep.*, 1960, **6**, 339–345.

Goodrich, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–63.

Hull, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Ramond, C. K. Performance in instrumental learning as a joint function of delay of reinforcement and time of deprivation. *J. exp. Psychol.*, 1954, **47**, 248–250.

Reynolds, W. F., & Pavlik, W. B. Running speed as a function of deprivation period and reward magnitude. *J. comp. physiol. Psychol.*, 1960, **53**, 615–618.

Spence, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In, *Learning theory, personality theory, and clinical research: The Kentucky symposium.* New York: Wiley, 1953. Pp. 1–21.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Spence, K. W. *Behavior theory and learning.* N. J.: Prentice Hall, 1960.

Wagner, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

Weiss, R. F. Deprivation and reward magnitude effects on speed throughout the goal gradiant. *J. exp. Psychol.*, 1960, **60**, 384–390.

(Received February 27, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 472–479

THE INDUCTION OF NONVERIDICAL SLANT AND THE PERCEPTION OF SHAPE[1]

WILLIAM EPSTEIN, HELEN BONTRAGER, AND JOHN PARK

University of Kansas

Koffka's (1935, pp. 224–235) explanation of shape-constancy is based on the assumption of an invariant linkage between slant and shape. Koffka's hypothesis has been formulated in the following way by Beck and Gibson (1955): "A retinal projection of a given form determines a unique relation of apparent shape to apparent slant" (p. 126). The invariance hypothesis requires that perceived shape vary as a strict function of variations in perceived slant (Langdon, 1951, 1953; Nelson & Bartley, 1956; Stavrianos, 1945).

In order to test this hypothesis Beck and Gibson (1955) induced errors in the apparent slant of a target whose shape was to be judged. A triangular target cut from untextured, white cardboard was mounted on a regularly textured, vertical background, and was slanted outward from the base at an angle of 45° from the background. The target was viewed monocularly with a motionless head. Under these conditions, in the absence of binocular disparity, motion perspective, and texture gradients, the slant of the target was misperceived. The triangle assumed the slant of the background, i.e., it appeared to be perpendicular to the line of sight instead of slanted 45°. Along with this standard target, two comparison triangles, placed flat on the background were exhibited. One represented a frontal parallel projection of the standard, and the other had the same physical dimensions as the standard. All *S*s judged the standard triangle to be equal in shape to the projective comparison. This match agrees perfectly with the requirements of the invariance hypothesis. When the same judgments were made with unrestricted binocular vision, 77% of the *S*s selected the objectively equal comparison triangle. This result also is in the direction required by the invariance hypothesis.[2] Since unrestricted observation eliminates the slant-inducing effect of the background, accurate perception of slant was restored; and, with it, veridical perception of shape was also recovered.

The general purpose of the present study was to extend and clarify Beck and Gibson's (1955) findings. The following were the major extensions and modifications: (*a*) The slant-induction effect was investigated when the background also was slanted from the frontal parallel plane. (*b*) A technique was employed which permitted continuous variation of the comparison stimulus. This made possible a more exact test of the invariance hypothesis. (*c*) A more accurate measure of the apparent slant of the target was obtained. These data are necessary for an analysis of the slant-shape relationship. (*d*) The influence of three different instructional sets on the judgment of shape was investigated. Beck and Gibson (1955) do not inform us concerning this variable, and the evidence from other studies (Gottheil & Bitterman, 1951; Klimpfinger, 1933) is fragmentary.

METHOD

Apparatus

The main apparatus was a rectangular light-tight tunnel 7 ft. in length with walls

[1] This investigation was supported by grants to the first author from the National Institute of Mental Health of the United States Public Health Service (M-4153) and the General Research Fund of the University of Kansas. We wish to acknowledge the assistance of Raymond Engstrand who, together with the second author, collected the data and performed the statistical computations.

[2] The invariance hypothesis actually requires a 100% shift to the objectively equal comparison.

20 in. wide. The interior of the tunnel was painted flat black. In the front wall of the tunnel there was an aperture which could be adjusted for monocular or binocular vision. A head-clamp chin-rest arrangement placed in front of the aperture kept *S*'s head motionless when this was required. The two clamps restrained all but the most determined head movements, and head movements are considered highly unlikely under the conditions of the experiment. A reduction tube was inserted in the front wall which restricted the monocular field to the target and its immediate background. On the inside of the front wall was a circular fluorescent lamp (32 w., standard cool white color) which provided the only illumination in the tunnel. This lighting arrangement made invisible the shadow cast by the slanted standard. At a distance of 6 ft. from the front wall a false back wall (panel) was inserted. This panel was covered with a black and white checkerboard cloth composed of 1×1 in. squares. The panel could be adjusted from outside the tunnel to three degrees of slant: perpendicular to the line of sight (vertical), 20° from the perpendicular tilted away from *S* (20° A), and 20° from the perpendicular tilted toward *S* (20° T).

The standard triangle was mounted in the objective center of the panel and directly in *S*'s line of regard, and had a height of 5 in. and a base of 4 in. The adjustable comparison triangle was mounted flat on the background, 4 in. above the apex of the standard. The height and base of the comparison could be varied continuously by manipulating a set of levers located on the exterior of the roof of the tunnel. The design of this comparison stimulus is similar in essentials to the apparatus described by Gottheil and Bitterman (1951, p. 407). The comparison and standard triangles were cut from the same white cardboard of imperceptible texture.

The comparison stimulus for slant was a circular disc of white cardboard mounted on a horizontal axis. The disc could be rotated on its axis, and its angle of rotation could be read directly off a protractor. The disc was located directly in *S*'s line of sight when he turned 90° into the designated viewing position, and it was presented outside the tunnel in a normally illuminated, unrestricted field.

Procedure

Stimulus conditions.—The standard was judged twice by each *S* (one ascending and one descending trial) for each of three background slants. When the background was vertical or 20° T, the standard triangle was at a 45° slant from the background outward from the base. When the background was slanted 20° A, the standard was slanted 45° outward from the apex.

Conditions of observation.—Each *S* observed the standard either monocularly or binocularly.

Instructional conditions.—Three sets of instructions were used which were intended to induce different attitudes of observation. These attitudes are usually designated: the phenomenal, objective, and analytic. Each *S* served under one instructional condition only. The instructions were as follows:

> (Instructions for the phenomenal attitude) When you look into this box you will see a standard triangular target on a checkerboard background. I am interested in learning how you perceive the shape of this target. Right above the standard triangle there is a second triangle. The base and altitude of this second triangle can be varied. Your task will be to instruct me to adjust the shape of the variable triangle so that it appears to be the same shape as the standard below it. I would like you to suspend all mental judgments and give me a match which reflects your immediate perceptual impression. Don't try to figure out a good match. I want a report of your immediate perceptual impression even if you feel that the match you make would not agree with the objective physical dimensions of the target.
>
> (Instructions for the objective attitude) I am interested in learning how well you can reproduce the actual physical shape of this target. . . . This means that in the ideal case when you have completed your judgment I should be able to take the match you have made, lay it over the standard, and find that it corresponds perfectly in all dimensions. Please remember that you are to reproduce the actual physical dimensions of the target even if the match you make doesn't *look* equal in shape to you. To make this clear suppose you were looking at a man far in the distance. He would *look* very small but if you were asked to reproduce the actual size of the man you probably would be fairly accurate. This is what I wish you to do here. Reproduce the actual physical shape of the target.
>
> (Instructions for the analytic attitude) I am interested in learning how well you can reproduce the retinal shape of this target. An illustration may help make this

clear. If I put a coin on the table and you look directly down upon the coin, it will project a circular retinal image. However, if you stand back along the edge of the table and look at the coin from this new position, the coin will produce an elliptical retinal image. It is this retinal shape in which I am interested. . . . Your task will be to instruct me to adjust the shape of this variable triangle so that it is identical in shape with the retinal shape of the standard triangle. Remember I am interested in the retinal shape. It is not important to me whether or not the two triangles look identical in shape or are actually equal in shape when you have completed your judgment.

The shape-matching procedure may now be summarized. Each *S* was instructed to make two shape-judgments of the standard for each of three background slants. The *S* made the judgments by instructing *E* to adjust the shape of the comparison triangle until a satisfactory match was attained. Each *S* performed this task under one observational condition, binocular or monocular, and under one instructional set, phenomenal, objective, or analytic.

When the shape-judgments were concluded, the standard was presented again, and *S* judged the slant of the standard once for each slant of the background. All *S*s were given the following instructions:

Now I would like you to judge the slant of the target triangle. By "slant" I am referring to whether the triangle appears to stand straight up and down in front of you, or whether it appears to lean either toward or away from you. This is how your slant judgments will be made. First, you will look at the triangle and determine if it is slanted and, if so, how much. Outside of the box I will show you a circular disc whose slant may be varied. Your task is to adjust the slant of this disc to the same slant as the target. After you have adjusted the slant of the disc you will be given one more look into the box to check your judgment.

The experiment was concluded with an interview which sought to obtain information about three aspects of the experiment: (*a*) the effectiveness and comprehensibility of the instructions; (*b*) the deliberate utilization of perceived slant during the judgment of shape; and (*c*) the relation between the judgments of slant obtained *after* the shape-judgments were completed and the perceived slant *during* the shape-judgments.[3]

Subjects

The *S*s were 84 undergraduates who served in the experiment as a course requirement. They were assigned in order of their appearance to one of six groups (2 conditions of observation × 3 attitudes of observation). Thus, each of the six experimental groups contained 14 *S*s, about equally divided among men and women. None of the *S*s knew anything about the invariance hypothesis, or the questions under investigation.

Results

The main results of the experiment are presented in Table 1. In all of the tables the shape data are represented as height-base (h/b) ratios.

Slant-induction.—An inspection of the mean slant judgments for monocular observation recorded in Table 1 reveals that the slant-inducing effect of the background reported by Beck and Gibson (1955) was obtained for all three background slants. As an illustration consider the fourth row of figures in Table 1. When the background was 20° A and the standard was objectively slanted 65° A, the standard was judged to be slanted 17.21° A. When the background was vertical and the standard 45° T, the standard was judged to be slanted 0.83° T. In the case of the 20° T background slant, an exaggerated slant-induction effect appears to have occurred. While the standard was objectively slanted 65° T, it was perceived to be at a slant of 6.59° T, i.e., slanted 13.41° less than the background.

The comparable data for binocular vision show the expected diminution of the slant-induction effect when the

[3] The use of successive rather than simultaneous slant-shape judgments can be questioned. We sought to eliminate some of the ambiguity during the interview.

TABLE 1

MEAN h/b RATIOS OF COMPARISON MATCHES AND MEAN SLANT JUDGMENTS OF STANDARD

Observation Attitude	Slant of the Background											
	20° A				Vertical				20° T			
	Comparison		Standard		Comparison		Standard		Comparison		Standard	
	Mean h/b	*SD*	Mean Slant	*SD*	Mean h/b	*SD*	Mean Slant	*SD*	Mean h/b	*SD*	Mean Slant	*SD*
	Monocular Vision											
Phenomenal	.370	.043	10.57° A	11.40°	.573	.053	0.80° T	3.37°	.386	.063	4.06° T	7.58°
Objective	.397	.018	17.30° A	12.00°	.556	.043	0.85° T	3.58°	.386	.044	10.21° T	6.48°
Analytic	.387	.097	23.57° A	13.96°	.530	.090	1.15° A	4.75°	.392	.091	7.64° T	7.54°
All attitudes	.384	.127	17.21° A	13.03°	.553	.180	0.83° T	4.43°	.388	.128	6.59° T	7.49°
	Binocular Vision											
Phenomenal	.547	.177	55.72° A	9.57°	.649	.077	20.57° T	8.89°	.584	.117	25.00° T	10.63°
Objective	.706	.102	52.72° A	14.07°	.710	.095	17.57° T	9.45°	.666	.129	35.28° T	12.53°
Analytic	.612	.153	56.65° A	7.91°	.651	.079	20.35° T	6.74°	.507	.134	29.35° T	11.27°
All attitudes	.625	.202	55.03° A	11.40°	.696	.213	19.49° T	7.28°	.585	.192	29.87° T	12.36°
Objective h/b (St.)	.800				.800				.800			
Proj. h/b (St.)	.353				.585				.324			
Objective slant (St.)[a]	65° A				45° T				65° T			

[a] Measured in terms of the deviation from the vertical away from (A) or toward (T) *S*.

cues for slant were restored. The case of the 20° A background illustrates this diminution. When viewed monocularly, the target appeared to be at a slant of 17.21° A—a deviation of 47.79° from its objective slant, but only 2.79° from the slant of the background. However, when viewed binocularly, the target was judged to be at a slant of 55.03° A—a deviation of only 9.97° from its objective slant, but 35.03° from the slant of the background.

In order to assess the relative magnitude of the slant-induction effect, the slant judgments were expressed as deviations from the slant of the background, and an analysis of variance was performed on these deviation scores. The analysis confirmed the observations recorded above. For any given degree of background slant the deviation was significantly greater for binocular than for monocular vision, i.e., the slant-induction effect was greater for monocular vision. The magnitudes of the deviations were not equal for all background slants within each condition of observation. The 20° T background slant produced the greatest deviation for monocular vision and the smallest deviation for binocular vision. However, if the large deviation for monocular vision may be assumed to be an exaggerated slant-induction effect, then we can conclude that the induction effect was more pronounced when both target and background were tilted forward.

Instructional sets.—An analysis of variance showed a significant interaction of attitude with condition of observation. When the target was viewed monocularly, instructional sets were ineffectual, i.e., they did not influence the shape-judgments. However, when the target was observed

TABLE 2

THE MEAN PROJECTIVE h/b RATIO REQUIRED BY THE APPARENT SLANT OF THE STANDARD COMPARED WITH THE MEAN PROJECTIVE h/b RATIO OF THE COMPARISON SETTING

Observation Attitude	Monocular			Binocular		
	Co. Mean Proj. h/b Ratio	St. Mean Proj. h/b Ratio	Sign and % Deviation	Co. Mean Proj. h/b Ratio	St. Mean Proj. h/b Ratio	Sign and % Deviation
	Background Slant = 20° A					
Phenomenal	.347	.200	+ 73.50	.507	.674	−24.77
Objective	.375	.251	+ 49.40	.656	.674	− 2.67
Analytic	.362	.343	+ 5.84	.541	.683	−20.79
	Background Slant = Vertical					
Phenomenal	.573	.609	− 5.74	.649	.735	−11.56
Objective	.556	.601	− 7.65	.710	.732	− 3.14
Analytic	.529	.608	− 12.17	.631	.733	−13.50
	Background Slant = 20° T					
Phenomenal	.368	.135	+172.59	.562	.409	+37.89
Objective	.367	.223	+ 64.12	.638	.514	+24.12
Analytic	.375	.208	+ 79.80	.483	.460	+ 5.00

binocularly, there were significant differences between the shape-judgments made under the various instructional sets. Here we will mention only that the objective attitude consistently resulted in matches which were significantly nearer to the objective dimensions of the target than the matches for the analytic and phenomenal attitudes. This finding is in agreement with the results reported by earlier investigators (Gottheil & Bitterman, 1951; Klimpfinger, 1933).

TABLE 3

SELECTED SAMPLE OF DATA FROM *S*s WHO DISPLAYED COMPLETE SLANT-INDUCTION WITH VERTICAL BACKGROUND AND MONOCULAR VISION

Observation Attitude	*N*	St. Objective h/b Ratio	Co. Mean Proj. h/b Ratio	St. Proj. h/b Ratio	Sign and % Deviation
Objective	10	.800	.540	.585	−7.68%
Analytic	9	.800	.543	.585	−7.17%
Phenomenal	7	.800	.555	.585	−5.12%
All attitudes	26	.800	.546	.585	−7.85%

Apparent slant-apparent shape invariance.—The following procedure was used to evaluate the invariance hypothesis. An h/b ratio was determined trigonometrically for each *S*'s *apparent* (judged) slant of the standard at each background slant. This ratio represents the projective ratio of the shape-match demanded for that degree of apparent slant by strict adherence to the invariance requirements. Next, the projective h/b ratio was determined for *S*'s setting of the comparison triangle at each background slant. A comparison of these two ratios allowed us to evaluate the degree of slant-shape invariance. Any difference between the two ratios represents a deviation from invariance requirements. The

results of our analysis are summarized in Table 2.

To clarify the logic of this analysis a selected sample of data has been presented in Table 3. The mean h/b ratios contained in Table 3 were obtained from *Ss* who manifested total slant induction when viewing the standard monocularly against a vertical background. In this special case, when the standard is perceived to be perpendicular to the line of sight, the frontal parallel projection for the apparent slant of the target is in effect the actual projective shape of the target. This is to say that, if the invariance requirements are to be strictly satisfied, the projective h/b ratio of the comparison (which in this special case is identical with the objective h/b ratio of the comparison) must equal the projective h/b ratio of the standard. For the dimensions of our standard this means that a perfect invariance fit would be h/b = 0.585. However, consulting the fourth column in Table 3, we find that this theoretical value was not attained. The deviations between the means were not great, but they were present, nonetheless.

An inspection of Table 2 reveals that in no instance did the mean projective h/b ratio of the comparison setting satisfy the invariance requirements. The deviations obtained for monocular observation ranged from −5.74% to 172.59%. For binocular vision the range of deviations was −2.67% to 37.89%. Relatively small deviations from invariance were obtained when the background was vertical. In addition, there was no significant difference between the grand mean deviations for monocular and binocular vision with the vertical background. The greatest deviations occurred when the background was slanted 20° T, although errors of considerable magnitude were obtained with the background 20° A. With only one exception, the deviations obtained with monocular observation under the 20° A and 20° T background conditions were at least twice as great as the deviations obtained for binocular vision.

A review of the ratio pairings for the individual *Ss* (252 pairings, 126 for monocular and the same number for binocular vision) confirms these observations. There was not one case of perfect correspondence between the two h/b ratios. Table 4, which presents the distribution of deviations according to sign, reflects this fact.

Verbal reports.—The information derived from the interview tended to confirm the validity of the experimental operations. All *Ss* gave evidence of having understood the task which was posed by the instruc-

TABLE 4

DISTRIBUTION OF DEVIATIONS FROM INVARIANCE REQUIREMENTS FOR MONOCULAR AND BINOCULAR VISION

Observation Attitude	Slant of Background						All Slants Deviations	
	20° A Deviations		Vertical Deviations		20° T Deviations			
	+	−	+	−	+	−	+	−
Monocular Vision								
Phenomenal	12	2	3	11	13	1	28	14
Objective	10	4	2	12	12	2	24	18
Analytic	8	6	3	11	13	1	24	18
All attitudes	30	12	8	34	38	4	76	50
Binocular Vision								
Phenomenal	3	11	3	11	11	3	17	25
Objective	6	8	8	6	11	3	25	17
Analytic	3	11	2	12	8	6	13	29
All attitudes	12	30	13	29	30	12	55	71

tions. Very few *S*s said that they had made deliberate efforts to take slant into account while making shape-judgments. All *S*s agreed that the judgments of slant made at the end of the experiment corresponded well with the apparent slant of the target during the earlier judgments of shape. Only a few *S*s expressed any strong lack of confidence in their judgments.

Discussion

The results of this experiment confirmed and extended Beck and Gibson's (1955) findings concerning the slant-induction effect. The effect was observed to obtain not only when the background was vertical, but also when the background was slanted 20° A or 20° T. In fact, there is some reason for concluding that the effect was more pronounced when the background was slanted forward. We have no explanation for this latter finding, nor do we see any way of deriving it from the general stimulus-gradient theory.

The results with regard to the invariance hypothesis agreed with Beck and Gibson's findings in their general aspects, but were sufficiently different to warrant comment. The main point of disagreement was this: For both monocular and binocular observation our results showed much less adherence to the invariance requirements than did the results of Beck and Gibson. For instance, Beck and Gibson report that when the standard was viewed monocularly with a stationary head, all of their 30 *S*s selected a projective match, while with unrestricted binocular vision 23 of 30 *S*s selected an objective match. None of our *S*s made a perfect projective or objective match when these were demanded by the apparent slant.

The main reason for this disagreement resides in the difference between the measurement techniques employed in the two experiments.[4] We used a continuously variable comparison which enabled *S* to make sensitive discriminations. The method which Beck and Gibson used forced *S* to choose between one of two extreme alternatives, i.e., objective or projective. There was no way *S* could indicate any other perceived shape. In this situation *S* would probably select the comparison triangle which was *most like* the one he perceived even when he recognized that the two stimuli were not identical. Thus, all perceived shapes which clustered about the projective comparison were designated by that comparison and, similarly, all apparent shapes which clustered about the objective comparison were designated by that comparison. Lacking an opportunity to differentiate in the response system, *S*s produced results which appeared to reflect agreement in the perceptual system.

The evidence from past studies with regard to the effect of attitudes on the perception of shape is difficult to interpret. Klimpfinger (1933), who is most frequently cited, did not actually compare the influence of different attitudes within the same experimental situation. Instead, his conclusions were based mainly on a comparison of his data with the results of other *E*s. Gottheil and Bitterman (1951) used the same *S*s for each of the three attitudes. There is no way of knowing what effect judgments made under a previous attitude may have had on the later judgments obtained for another attitude. In addition, Gottheil and Bitterman do not tell us whether *S* viewed the target monocularly or binocularly. Our own results suggest an interaction of attitude with condition of observation. In light of the paucity of evidence on this question further research is needed.

Summary

The experiment had three aspects: (*a*) to determine whether the slant-induction effect reported by Beck and Gibson (1955) would obtain when the background was slanted; (*b*) to test more precisely the slant-shape invariance hypothesis; and (*c*) to investigate the influence of three attitudes of

[4] Beck and Gibson (1955, p. 131) recognized the limitations of their technique for a precise test of the invariance hypothesis.

observation on the perception of shape under monocular and binocular vision.

The main results were: (*a*) The slant-inducing effect of the background was observed when the background was slanted 20° from the perpendicular either away from *S* (20° A) or toward *S* (20° T) in addition to the case in which the background was perpendicular to the line of sight. (*b*) Under no condition did the judgments of apparent slant and apparent shape covary exactly as demanded by the invariance hypothesis. The deviations from invariance ranged from −24.77% to 172.59%. (*c*) The influence of attitudes on the perception of shape was restricted to binocular viewing. Attitudes were found to be ineffectual in determining perceived shape when the standard was viewed monocularly.

References

Beck, J., & Gibson, J. J. The relation of apparent shape to apparent slant in the perception of objects. *J. exp. Psychol.*, 1955, **50**, 125–133.

Gottheil, E., & Bitterman, M. E. The measurement of shape-constancy. *Amer. J. Psychol.*, 1951, **64**, 406–408.

Klimpfinger, S. Über den Einfluss von intentionaler Einstellung und Übung auf die Gestaltkonstanz. *Arch. ges. Psychol.*, 1933, **88**, 551–598.

Koffka, K. *Principles of Gestalt psychology.* New York: Harcourt, Brace, 1935.

Langdon, F. J. The perception of changing shape. *Quart. J. exp. Psychol.*, 1951, **3**, 157–165.

Langdon, F. J. Further studies in the perception of a changing shape. *Quart. J. exp. Psychol.*, 1953, **5**, 89–107.

Nelson, T. M., & Bartley, S. H. The perception of form in an unstructured field. *J. gen. Psychol.*, 1956, **54**, 57–63.

Stavrianos, B. K. The relation of shape-perception to explicit judgments of inclination. *Arch. Psychol., N. Y.*, 1945, No. 296.

(Received March 20, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 480–486

RETENTION OF CONNECTED DISCOURSE AS A FUNCTION OF DURATION OF INTERPOLATED LEARNING[1]

NORMAN J. SLAMECKA

University of Vermont

Several verbal learning studies have shown that recall of original learning (OL) varies inversely as a function of the degree of interpolated learning (IL) (Slamecka & Ceraso, 1960, p. 451). By *degree* of IL is meant the number of trials of presentation of a single IL list. Further, it is also known that recall of OL varies inversely with the amount of IL (Twining, 1940; Underwood, 1945). By *amount* of IL is mean the total number of different IL items presented. Amount of IL can be manipulated either by varying the number of different IL lists, or by varying the length of a single IL list.

The question whether amount of IL or degree of IL produces more forgetting has definite empirical value and theoretical interest, but to date there have been only two attempts to investigate it, and they both exhibit certain limitations. Underwood (1945) manipulated amount by varying the number of different IL lists. Using paired associates in an A-B, A-C paradigm, he compared three levels of degree of IL (8, 16, and 24 trials on a single list) with three levels of amount of IL (2, 4, and 6 different lists, with 4 trials on each list), in such a way that each level was equated for total trials. Results showed that retention decreased at a faster rate with increases in amount than with increases in degree of IL. However, it is rather ironic that in this unique experiment the usual progressive decrements in recall with increasing degrees of IL, which are commonly reported in other studies, were not obtained, and therefore the generalizability of the results is questionable. Had the expected progressive effectiveness of the degree variable been in fact obtained, the conclusions might have been quite different. As it is, the conclusion that with paired associates, amount produces more forgetting than degree must be regarded as tentative. McGeoch (1936), using serial lists of adjectives, reported the only study in which amount was manipulated by varying the length of the IL list. Interpolated lists were 8 or 16 units long, with presentation frequencies of 4, 8, and 16 trials. As in the Underwood study, differences in degree of IL within either of the list lengths failed to produce significant differences in recall. Further, although it was found that an 8-unit list given 8 or 16 trials produced slightly less retroactive inhibition (RI) than a 16-unit list given 4 or 8 trials, the differences were small and not statistically reliable. It is felt, therefore, that the question of the effects upon retention of these various ways of manipulating IL is still in need of investigation.

The present paper reports a series of three related RI experiments which seek to determine whether there is a common factor responsible for interference in the rote retention of certain serial materials, so designed as to allow direct cross comparisons among the experiments. Experiment I varied the degree of IL, Exp. II varied the amount by using different numbers of IL passages, and Exp. III used passages twice the length of those in the other two experiments. It was hypothesized that rote retention would be inversely related both to degree of repetition and to the number of IL passages given. Experiment III introduced a change in total trials across levels, since Exp. I and II were both comparable at each level in that regard. No hypotheses were formulated about any of the across-experi-

[1] This study was supported by National Science Foundation Grants G-6192 and G-14,721. The author is gratefully indebted to John A. Corson, who collected the data.

ment comparisons; however, the main intent and interest of the study is ultimately to focus upon those comparisons.

EXPERIMENT I

Method

Subjects.—Twenty-four undergraduate students of introductory psychology served as *Ss* as part of the course requirements. Each *S* had five individual sessions, the first of which was a familiarization period designed to acquaint *S* with memory drum procedure and to lessen practice effects. This practice session involved the learning of a 20-word passage of connected discourse to one perfect trial, followed by six trials on an IL passage, and then two OL relearning trials. Each *S* was tested under all four experimental conditions.

Materials.—Serial materials consisting of 20-word passages of connected discourse were used. The following OL passages were presented:

1. Any generalized statement of relationship between variables requires specification of some reference class or universe from which the material comes.

2. According to recent considerations it appears preferable to select preliminary assumptions from the realm of objectively verifiable and observable phenomena.

3. Traditional rationalism cannot negate the assertion that its theory concerning primary evidence rests upon inadequate analysis of fundamental scientific methodology.

4. Forthright admission of the contingent character of our most soundly established beliefs distinguishes contemporary empiricists from dogmatic and positive ancients.

The following IL passages were used, of which each *S* was given three:

1. Early logical writings formulated the rationalist ideal of knowledge, but exploratory biological works exhibit less exacting standards of systematic inquiry.

2. Primitive conceptions of organization, such as classical geometry, were believed to be based upon completely demonstrable and absolutely indubitable principles.

3. Current convictions maintain that methods of natural science are the most reliable instruments thus far devised for comprehending empirical relations.

4. Next to establishment of the peripheral anchors of behavior, the observational tracing of mediating physiological processes has also merited investigation.

5. In the existing systems of measurement no contemporary laboratory investigator will state conclusions without stipulating their coefficient of probable error.

6. Psychology must insist upon greater precision than even the physicist may find necessary in his routine description of experimental situations.

7. The derivative modes of expression are more frequent in biology than in other natural disciplines, however, they definitely facilitate conceptualization.

8. The language of classical physics in the context of subatomic research is valuable in focusing attention directly upon important analogies.

9. Science rests upon assumptions, since every conclusion presupposes premises which are either primary or establishable by reliance upon other propositions.

10. These definitions contain all the necessary conceptual elements for making explicit the general characteristics of an objective approach to psychology.

11. Our knowledge is not infallible, but our general experience is evidence that even inconclusive formulations may differ in their adequacy.

12. Certain crucial expressions currently applied to quantum theory denote different traits of things than when used in discussing earlier formulations.

Design and procedure.—There were four degrees of IL: 0, 6, 12, and 18 trials, to be called, respectively, the Control, Low, Medium, and High. The standard RI design was used, consisting of OL (to a criterion of one perfect trial), IL, and relearning (RL) for 4 trials. The passages were presented on a memory drum at a 3-sec. presentation rate and a 6-sec. intertrial interval. OL passages and IL conditions were counterbalanced with respect to *Ss* and sessions. IL passages were counterbalanced with respect to *Ss* and conditions. The time from the end of OL to the start of RL was 21 min. The IL trials began 30 sec. after the end of OL, leaving, in the case of the Control, Low, and Medium conditions, periods of about 21, 14, and 7 min. free of any learning task. This time was filled by having *S* judge pictures from the Meier Art Judgment Test. To compensate for a loss in the set to respond to the memory drum, the above three groups also had one trial of number guessing for warmup just preceding RL.

Results

Mean OL trials to criterion in Exp. I (including the criterional trial) are found in Table 1. There were no significant differences among the means ($F = 1.14$, $df = 3/69$). Since interpolated trials were fixed, the IL acquisition is presented in terms of mean total correct anticipations, also shown in Table 1. It is evident that considerable differences were produced between the levels. Retention data for the first two RL trials in Exp. I are found in Table 2. By RL Trial 3 almost all *S*s were at a uniformly high level of relearning. Analysis of recall (RL Trial 1) showed a significant effect of degree of IL in the predicted direction ($F = 16.35$, $df = 3/69$, $P < .01$), and t tests between adjacent means showed significant differences between the Medium and Low IL conditions ($P < .05$, $df = 69$), and the Low IL and Control ($P < .01$, $df = 69$) conditions. These results support the experimental hypothesis and lend further corroboration to the findings of a previous study of degree of IL in connected discourse, wherein different acquisition criteria and different passages were used (Slamecka, 1960).

Experiment II

Method

Subjects.—Another 24 students of introductory psychology served as *S*s. As in Exp. I, each took part in five sessions, the first of which was a practice period. Each *S* had all four experimental conditions.[2]

Materials.—The same passages as those used in Exp. I were used in OL and in IL.

Design and procedure.—This experiment varied the number of different IL passages presented. The four levels used were zero, two, four, and six passages (to be called, respectively, Control, Low, Medium, and High), each being presented for three consecutive trials. The RI design was followed, with the same OL criteria, rate of presentation, intertrial interval, and OL-RL interval, warmup trial, filler task, and the four RL trials as in Exp. I. When the third trial on an IL passage was completed, *E* then exposed a new passage (after the normal intertrial interval) with no loss of tempo. The OL passages and IL conditions were counterbalanced with respect to *S*s and sessions. The IL passages were counterbalanced so that each one appeared an equal number of times within each condition.

[2] Two *S*s had to be replaced because of their inability to keep to the appointment schedule.

Results

Mean OL trials to criterion for Exp. II are shown in Table 1. There were no significant differences among the means ($F = .66$, $df = 3/69$). Mean total correct anticipations during IL, also in Table 1, show obvious differences. Analysis of the recall means, found in Table 2, showed a significant effect of amount of IL ($F = 23.10$, $df = 3/69$, $P < .01$), and t tests indicated that all adjacent means differed significantly from each other in the predicted direction ($P < .01$ in all cases except Medium vs. High, where $P = .05$, $df = 69$). Therefore, the hypothesis that the retention of connected discourse is inversely related to the number of IL passages presented was given confirmation.

Experiment III

Method

Subjects.—Twenty-four students drawn from introductory psychology classes served as *S*s for five sessions, as previously described.

Materials.—The same OL and IL passages as in Exp. I and II were used. The only change was that the length of a given IL passage was doubled by linking two of the 20-word passages so as to form a two-sentence unit of 40 words to be learned. Thus, one trial with such a composite passage was equivalent, in elapsed time, to two trials with the shorter passages used previously. A total of six composite passages was obtained by combining the following numbered short IL passages: 1–7, 2–11, 3–4, 9–5, 10–6, 12–8.

At each session, only one such passage was presented.

Design and procedure.—There were four levels of interference, designated by the number of trials given on a single composite passage. These were 0, 3, 6, and 9 trials, to be called, respectively, the Control, Low, Medium, and High conditions. It should be noted that the total presentation time for these passages, at each level, was equivalent to that in the other two experiments, since their increased length was compensated by halving the total trials. Essentially, this was a variant of the degree of IL study, and therefore the same experimental hypothesis would apply. The OL passages were counterbalanced with respect to sessions, *S*s, and conditions. The composite IL passages were counterbalanced with sessions, *S*s, and OL passages. Although complete counterbalancing of IL passages with conditions was not possible, a reasonable approximation to it was achieved. The general procedure followed was the same as in Exp. I and II.

Results

Mean OL trials for Exp. III are shown in Table 1. There were no significant differences among the means ($F = .99$, $df = 3/69$). Mean total correct anticipations during IL, in Table 1, show large differences. Mean amount recalled, as shown in Table 2, reflected significant differences attributable to degree of IL ($F = 16.64$, $df = 3/69$, $P < .01$), and t tests between adjacent means showed significant differences beyond the .01 level between the Low and Medium, and between the Medium and High conditions. Thus, degree of IL with the lengthened passages was also an effective variable. The significance of all three experiments when considered together will be discussed below.

TABLE 1

MEAN OL TRIALS TO CRITERION AND MEAN TOTAL CORRECT ANTICIPATIONS IN IL FOR ALL THREE EXPERIMENTS

Exp.	Control		Interpolated Learning					
			Low		Medium		High	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
	OL Trials							
I	10.6	3.4	9.5	3.1	10.8	3.0	11.0	3.7
II	10.4	5.2	10.9	5.2	10.2	3.8	11.5	5.4
III	10.1	3.8	11.3	3.6	10.5	4.7	11.5	5.0
	IL Anticipations							
I			70.3	9.5	188.4	15.2	304.3	21.4
II			40.0	12.4	77.2	21.8	121.7	29.5
III			29.1	9.7	96.7	26.8	195.0	42.6

TABLE 2

MEAN RETENTION SCORES FOR THE FIRST TWO RELEARNING TRIALS FOR ALL THREE EXPERIMENTS

Exp.	Control		Interpolated Learning					
			Low		Medium		High	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
	First RL Trial							
I	17.8	2.0	15.3	3.1	13.9	3.4	12.7	3.3
II	18.0	1.4	16.2	2.3	13.0	4.1	11.6	3.6
III	17.3	2.1	16.2	2.8	13.8	3.7	11.7	3.2
	Second RL Trial							
I	19.4	0.6	18.8	1.6	18.4	1.5	18.2	2.4
II	19.2	1.5	19.0	1.3	17.8	2.0	17.6	2.1
III	19.1	1.1	18.8	1.6	17.7	2.5	17.1	2.0

INTEREXPERIMENTAL COMPARISONS AND DISCUSSION

The reason for conducting three comparable experiments was to enable more precise information to be obtained regarding the specific source of the interference. Hitherto, verbal learning studies of the RI type have suffered from a form of confounding when seeking answers to this question. For example, the typical study of degree of IL employs varying numbers of trials on a single IL list. But, variation in the number of trials also results in concomitant variation in the total correct anticipations and in the total duration of the interfering learning as well. By the same token, a study of amount of IL in which the number of different lists is varied also involves concomitant variation in total IL time and total number of

different IL items presented (see for example, Twining, 1940, where number of lists and number of items are completely confounded). This confounding can be circumvented by carrying out multiple, but nevertheless comparable experiments. The present study was designed so as to permit meaningful cross comparisons at each level, which would enable clearer identification of the factor or factors responsible for the forgetting.

At each of the three levels of interference there were wide differences in total trials, trials per passage, number of passages, number of different items, and mean total correct anticipations. Referring to Exp. I, II, and III in that order, we find that: (*a*) total trials were 6, 6, and 3 for Low; 12, 12, and 6 for Medium; and 18, 18, and 9 for High; (*b*) trials per passage were 6, 3, and 3 for Low; 12, 3, and 6 for Medium; and 18, 3, and 9 for High; (*c*) number of passages were 1, 2, and 1 for Low; 1, 4, and 1 for Medium; and 1, 6, and 1 for High; (*d*) number of different items were 20, 40, and 40 for Low; 20, 80, and 40 for Medium; and 20, 120, and 40 for High; and (*e*) mean total correct anticipations were 70, 40, and 29 for Low; 188, 77, and 97 for Medium; and 304, 122, and 195 for High. However, the one factor remaining constant within each level across all three experiments was the duration of IL, or the total time devoted to the interfering learning task. For the Control condition this was 0 min., for Low it was 6.6 min., for Medium it was 13.2 min., and for High, 19.8 min.

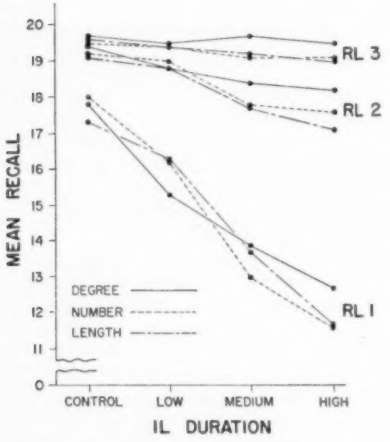

FIG. 1. Retention as a function of the time spent on the interfering learning task. (Control = 0 min., Low = 6.6 min., Medium = 13.2 min., and High = 19.8 min.)

The retention data have been plotted in Fig. 1 as a function of the IL durations. The Control conditions, where all *S*s had identical treatments, produced very similar recalls, testifying to the relative homogeneity of the *S*s used as well as to the degree of reliability of the techniques employed.

Analyses of the three recall (RL_1) means within each level gave *F*s of less than 1 in each case ($df = 2/69$), nor were there any within-level recall differences at the other relearning trials. Thus, equivalent amounts of forgetting were produced at each level by all three ways of manipulating the IL task. It made no difference how well the interfering items were learned, nor how many interfering items were presented, just as long as equivalent amounts of time were spent on the task. When forgetting was measured in terms of relative RI $\left[\frac{(\text{Rest-Work})}{\text{Rest}} \times 100\right]$, the percentages were: 14.0, 10.0, and 6.4 for Low; 21.9, 27.8, and 20.2 for Medium; and 28.6, 35.6, and 32.4 for High, indicating at least no consistent superiority of any one method over the others.

By virtue of the counterbalanced designs employed, the interference must be regarded as involving both RI and proactive inhibition (PI) together, since recent work has demonstrated that connected discourse is also subject to PI effects through the use of practiced *S*s (Slamecka, 1961). It might then be asked whether a similar pattern of retention scores would have been obtained had *S*s been naive. A partial approach to an answer with the present data would be to determine recall scores for performances on the first experi-

mental day, thereby using *Ss* with only the practice session as a possible source of PI. This results in a considerable loss of precision, since three-quarters of the original recall data would be unused. However, when calculated in this fashion, the recall means for Exp. I, II, and III, in that order, were: 19.0, 17.8, and 18.7 for Control; 14.3, 16.0, and 16.3 for Low; 15.3, 12.2, and 15.2 for Medium; and 13.5, 13.5, and 13.7 for High. Thus, even with only 6 *Ss* per data point, there is no evidence of a consistent difference in recall for any one experiment over the others, and under conditions of maximal interference the recalls are practically identical. Our main results, therefore, do not appear to be specific only to highly practiced or PI-burdened *Ss*.

For purposes of analysis, errors at recall were classed either as omissions, intralist intrusions, or "other" intrusions. In view of the small number of errors comprising the other class, it did not seem justified to attempt a further subclassification of them. Omissions accounted for three-quarters of all errors, and within-list sources accounted for two-thirds of the intrusion errors. Both classes of intrusions tended to rise with increasing interference and to maintain a relatively stable ratio with respect to each other in all three of the work conditions. Furthermore, there was no apparent trend in these ratios across the experiments. The ratios of within-list to other intrusions, in terms of mean proportions across experiments, were: 3.2 for Control, 1.7 for Low, 1.9 for Medium, and 1.8 for High. The ratios of within-list to other intrusions in terms of means across conditions were: 1.7 for Exp. I, 2.1 for Exp. II, and 2.1 for Exp. III. It appears that intrusion analysis does not reveal any characteristic differences among the three experiments. The percentages of all overt responses made that were correct decreased from the Control condition to the High IL condition. The average percentages over the three experiments were: 94.7 for Control, 93.7 for Low, 91.2 for Medium, and 87.2 for High. There was no significant variation in these percentages between experiments.

It is likely that with connected discourse, even the errors committed would be shaped by the contextual constraints and the sequential dependencies in the language. Many of the other intrusions were synonyms of the correct word, or else were otherwise appropriate to the context of the given passage. The within-list intrusions appeared to be largely anticipatory, as if *S* had skipped over some of the words, and such errors were also usually contextually appropriate.

The above analyses of recall, relearning, errors, and accuracy percentages, imply a quite clearcut principle, namely, that the rote retention of passages of connected discourse subjected to interpolated interference is solely a function of the amount of time spent on the interfering task, regardless of how that task is structured. There is no a priori reason why these results should not be as readily obtainable with serial lists of an unconnected or nonsense type, and if comparable findings are reported with such materials we would have a useful simplifying generalization about rote serial retention.

This study also has some bearing on explanations of forgetting which stress confusion of list membership, according to the notion that the more interfering material that is presented, the greater the opportunity for confusion among items from different lists. In Underwood's (1945) study, where the A-B, A-C list construction is designed for specific interference, such an hypothesis appears quite reasonable. But serial materials may not involve specific item interference, and thus the amount of interfering material would not necessarily result in changes in confusion of list membership. In the present study, the conditions of Exp. II should have maximized such confusion, since many IL passages were presented, each to such a low degree that differentiation would be quite weak. Nevertheless, retention was no less in Exp. II than in Exp. I, where only a single IL passage was given,

and for many more trials. Rather, it seems that the interference was nonspecific with regard to the form of the IL task, its magnitude determined by the time spent on that task, and the overt errors governed largely by local contextual constraints.

Summary

Three related RI experiments were made to determine more precisely the source of interference in the rote retention of passages of connected discourse. Experiment I varied the degree of IL by giving 0, 6, 12, or 18 trials on a single passage; Exp. II varied the number of different IL passages by presenting 0, 2, 4, or 6 passages, each for three consecutive trials; Exp. III used passages twice as long as those given in the other two experiments, for half the number of trials, viz., 0, 3, 6, and 9. It was hypothesized that retention would be inversely related both to the degree of learning and the number of IL passages presented. The main intent of the study, however, was an across-experiment comparison at each of the levels. The one common factor within each level was the IL duration which was, for each interference level, 0 min., 6.6 min., 13.2 min., and 19.8 min. The same passages were used in all experiments.

Results were as follows: (*a*) Exp. I and III verified the hypothesis that retention was inversely related to degree of IL; (*b*) Exp. II confirmed the hypothesis that retention was inversely related to amount of IL; (*c*) Cross-experiment comparisons at each level showed no significant differences at recall or relearning; (*d*) the pattern of errors at recall was rather comparable for all experiments, with most errors being omissions, and the bulk of intrusions coming from within the passage. Contextual constraints seemed to determine the kinds of overt errors made.

It was concluded that the rote retention of passages of connected discourse subjected to interpolated interference is solely a function of the amount of time spent on the interfering task, regardless of how that task is structured. The data failed to support notions stressing confusions of list membership at recall, but rather pointed to a nonspecific interference effect for serial materials.

References

McGeoch, J. A. Studies in retroactive inhibition. VII. Retroactive inhibition as a function of the length and frequency of presentation of the interpolated lists. *J. exp. Psychol.*, 1936, **19**, 674–693.

Slamecka, N. J. Retroactive inhibition of connected discourse as a function of practice level. *J. exp. Psychol.*, 1960, **59**, 104–108.

Slamecka, N. J. Proactive inhibition of connected discourse. *J. exp. Psychol.*, 1961, **62**, 295–301.

Slamecka, N. J., & Ceraso, J. Retroactive and proactive inhibition of verbal learning. *Psychol. Bull.*, 1960, **57**, 449–475.

Twining, P. E. The relative importance of intervening activity and lapse of time in the production of forgetting. *J. exp. Psychol.*, 1940, **26**, 483–501.

Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59**(3, Whole No. 273).

(Received March 20, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 487–494

EFFECTS OF INSTRUCTIONAL SET AND UCS INTENSITY ON THE LATENCY, PERCENTAGE, AND FORM OF THE EYELID RESPONSE [1]

I. GORMEZANO AND JOHN W. MOORE

Indiana University

The purpose of the present investigation was to compare the effects of voluntary and nonvoluntary blink instructions on the latency, frequency, and form of eyelid responses under classical conditioning and extinction procedures. The study was also concerned with determining the effects on these response variables of independently varying UCS intensity.

The procedure of the Iowa eyelid conditioning laboratory of rejecting *Ss* on the basis of response latency assumes there is a causal relationship between responses characterized by sharp closure, long duration, and short latency, and the voluntary attempt of *S* to blink to the CS (Spence, 1953; Spence & Ross, 1957, 1959; Spence & Taylor, 1951). Spence and Ross (1959) have attempted to provide analogous support for this assumption by maintaining that judged voluntary responses are ". . . similar in form to those given by *Ss* instructed to blink to the CS so as to avoid the air puff and by *Ss* who, following conditioning, reported they were blinking voluntarily in order to avoid the puff" (p. 377). (Presumably, *Ss* "avoiding" the air puff are attempting to attenuate the noxious effects of the UCS by having it fall on the closed eyelid.) Explicitly, Spence and Ross (1959) have stated ". . . the rationale behind this procedure of eliminating *Ss* whose responses were primarily of the short latency, voluntary form was that such responses were governed by different laws than those obtaining in the case of the longer latency class of responses" (p. 377). As yet, Hartman, Grant, and Ross (1960) have reported the only study which has attempted to examine the response characteristics of *Ss* instructed to blink to the CS to "avoid" the air puff. These investigators examined the latency distributions of *Ss* instructed to blink voluntarily under ready versus no ready signal and massed versus spaced trial conditions and reported qualitative correspondence with the latency distributions of judged voluntary *Ss* recovered from the data of Spence and Ross (1959) and Hartman and Ross (1961).

Since the latency distribution of judged voluntary *Ss* is itself recovered from eyelid conditioning data, the possibility exists that the latency distribution, and other response characteristics of judged voluntary *Ss*, would show some degree of correspondence with conditioning data obtained under a wide variety of independent variables. In the present investigation orthogonal comparisons were made of the response characteristics of voluntary blink instructed *Ss* with that of nonvoluntary blink instructed *Ss* (i.e., *Ss* given conventional "neutral" instructions). Further, the present investigation attempted to determine if variations in UCS in-

[1] This research was supported in part by a grant from the National Institute of Mental Health, Grant M-4034(A) and Grant G 16030 from the National Science Foundation. A brief report of the experiment was presented at the Psychonomic Society, Chicago, September 1960.

tensity had any differential effects on the response characteristics of *Ss* under the two instructional sets.

Method

Apparatus.—The apparatus was the same as that described in detail by Moore and Gormezano (1961). The *S*'s room was sound treated and contained two identical enclosures which permitted 1 or 2 *Ss* to be run in a session with identical stimulating and recording components. The *S* sat with his head comfortably positioned on a chin rest and facing into the enclosure. About 3 ft. in front of his eyes was a 10-cm. circular milk glass window. An abrupt change in the illumination of this window, from .19 to .26 mL., served as the CS. Also present in the enclosure was an intercommunication set and two 15-w. incandescent lamps. The *S* wore an elastic headband supporting an air-jet with a 2-mm. diameter orifice and a microtorque potentiometer used to pick up movement from his right eyelid. The signal from the potentiometer was amplified by a Brush dc amplifier and recorded by means of an ink-writing Brush penmotor. Brush penmotors also recorded the onset and termination of the CS and UCS. A Western Union tape transmitter controlled the intertrial intervals and activated electronic timers controlling the durations of the CS and UCS, and the CS-UCS interval. The duration of the CS was 600 msec. and the CS-UCS interval 500 msec. The UCS was a puff of compressed nitrogen delivered to the right cornea and terminated with the offset of the CS. The intensities of the air puffs, measured at their point of delivery to the eye, were sufficient to support a 50-, 100-, or 200-mm. column of mercury.

Design.—Twenty-four *Ss* were randomly assigned to each of the six cells of a 2×3 factorial design in which voluntary (V) and nonvoluntary (NV) blink instructions were made orthogonal to three UCS intensities of 50, 100, and 200 mm. The only restriction upon the random assignment of *Ss* was that the proportion of women to men in each of the six cells remain the same (8 women: 16 men). The two recording systems were also made orthogonal to the instruction and UCS intensity dimensions so as to remove their contributions to the error variance.

Procedure.—Instructions appropriate to *S*'s group were given over the intercommunication set. The instructions given to the nonvoluntary blink groups (NV50, NV100, and NV200) were as follows:

> Please listen carefully to the following instructions. Remain seated comfortably and keep looking at the milk-white disk in front of you. Do not touch anything on your head at anytime during the experiment.
>
> You will see and feel a series of stimuli during the experimental session. The stimuli will consist of a light and a puff of air. Be careful not to control voluntarily your natural reactions to the stimuli. Do not try to figure out the experiment. Keep as detached an attitude as possible and simply let your reactions take care of themselves.
>
> You can communicate with me at anytime by speaking in a normal voice. Are these instructions perfectly clear to you? Please tell me what you are to do.

The only difference in the instructions given to *Ss* in the voluntary blink instructed groups (V50, V100, and V200) occurred in the second paragraph which was given as follows:

> You will see and feel a series of stimuli during the experimental session. The stimuli will consist of a light and a puff of air. Your task is to blink so as to "avoid" the puff of air.

Following the instructions all *Ss* received 70 acquisition and 20 extinction trials. The intertrial interval for the 90 trials was randomized at 15, 25, and 35 sec. with a mean of 25 sec. The response characteristics of eyelid closures of at least 1-mm. deflection from the baseline, in the latency range of 150 to 525 msec. during acquisition and 150 to 600 msec. during extinction, were those subjected to statistical analyses.

Subjects.—The *Ss* were 96 men and 48 women from introductory psychology courses at Indiana University. In addition, 6 men and 10 women were discarded from the experiment because of apparatus failure or experimental error.

Results

The frequency distribution of response latencies in acquisition is shown in Fig. 1 for all six groups. The V groups are depicted on the left and the NV groups on the right with UCS intensity as the parameter. Each of the distributions were clearly bimodal. The responses in the first modes from the left were distributed

about a 75-msec. mode and are commonly taken as reflex responses to a visual CS (Grant, 1943). The low frequency of responses in the 125- to 200-msec. interval among the NV groups would seem to indicate that precautions to maintain light adaptation were successful in eliminating Beta responses (Grant & Norris, 1946). Concerning the principal modes appearing on the right, it is clear that instructional set had a considerable effect on response latency. The modal response occurred before 300 msec.

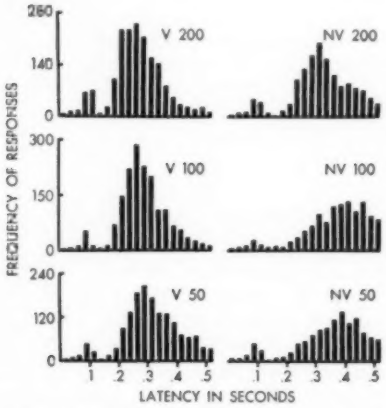

FIG. 1. Latency distribution of all eyelid responses in acquisition.

in the V groups and after 300 msec. in the NV groups. Further examination of the distributions indicates that as UCS intensity increased the modal response latency decreased under both instructional sets.

Figure 2 shows the frequency distribution of response latencies in extinction. Though all groups showed a drop in frequency of responses, the striking aspect of these distributions is that the modal responses of the V groups did not appear to have changed to any considerable extent from acquisition to extinction. On the other hand, the NV groups demonstrated particularly sharp decreases in response frequency around the modal response latency observed in acquisition (i.e., the 300- to 400-msec. interval).

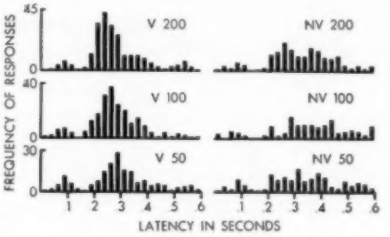

FIG. 2. Latency distribution of all eyelid responses in extinction.

Table 1 presents the effects of instructional set and UCS intensity on the latency and form of eyelid responses during acquisition and extinction. The cell entries for latency are the mean latency of responses in milliseconds and those for angle are the mean response angles in degrees. The response angle was determined from the graphic records by taking the angle formed between the baseline and the line passing through the points of response onset and maximum lid closure in the bounded interval of 150 to 525 msec. in acquisition

TABLE 1

MEAN LATENCY AND ANGLE OF EYELID RESPONSES DURING ACQUISITION AND EXTINCTION

Group	Latency (Msec.)		Angle (Deg.)	
	Acquisition	Extinction	Acquisition	Extinction
V50	331.2	331.6	28.3	25.9
V100	297.9	287.5	27.6	23.3
V200	284.8	288.6	27.8	23.8
NV50	384.3	399.6	20.3	21.9
NV100	390.7	400.0	17.7	17.2
NV200	351.6	358.0	19.9	20.6

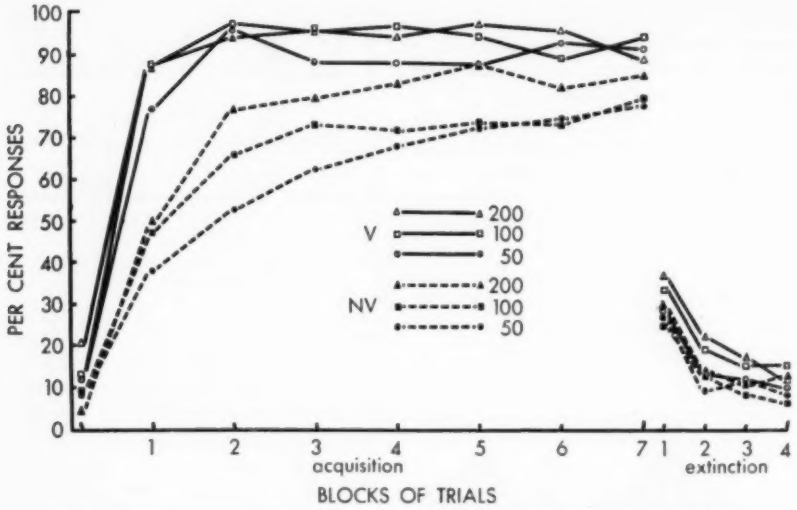

FIG. 3. The percentage of responses plotted in 10-trial blocks during acquisition and 5-trial blocks during extinction.

and 150 to 600 msec. in extinction. This measure represents an attempt to objectively quantify the form of eyelid closures. From examination of the table it is apparent that in acquisition and extinction, the V groups had considerably shorter mean latencies and sharper lid closures than the NV groups. It is also clear that as UCS intensity increased from 50 to 200 mm. the mean latency of responses decreased under both instructional sets, and response angle failed to show systematic change.

Figure 3 presents the performance curves in acquisition and extinction. The initial points for all acquisition curves are the mean percentage of responses on Trial 1 and the remaining acquisition points are the mean percentage of responses for each 10-trial block. The extinction curves are plotted in 5-trial blocks. In acquisition, the V groups proceeded to asymptote within 20 trials. Groups V100 and V200 stabilized at around 95% and Group V50 at about 90%. On the other hand, the curves for the NV groups rose gradually and the slopes of Groups NV50 and NV100 suggest that asymptotic performance for these groups had not been attained within the 70 conditioning trials. Group NV200 appears to have reached an asymptotic level of about 85%. The mean percentage of responses for Groups NV50, NV100, and NV200 in acquisition were 64.1, 69.0, and 78.0%, respectively, whereas the means were 88.4, 93.9, and 93.2% for Groups V50, V100, and V200, respectively. For the extinction curves presented on the right hand side of the figure, it is of particular interest to note that the higher level of responding of the V groups in acquisition also persisted in extinction. The mean percentage of responses in extinction for Groups NV50, NV100, and NV200 were 28.8, 27.1, and 34.2%, respectively, and for Groups V50, V100, and V200 the means were 32.5, 40.4,

and 45.4%, respectively. Thus, under each instructional set there was a small but increasing resistance to extinction as UCS intensity increased.

Statistical analysis of each of the response measures was carried out by analysis of variance. The latency and form analyses were performed on the log transforms of individual mean response latencies and individual mean response angles, respectively; the analysis of percentage of responses was performed on the arc-sine transform of individual percentage scores. These transformations served to bring about homogeneity of variance, making them more amenable to the analysis of variance (Bartlett, 1947). Table 2 gives the summary of the analysis for each of the response measures for acquisition and extinction. Except for the error mean squares which appear in parentheses at the bottom of each of the columns, only the F ratios are presented. In Row 1, the source of variation tested is the effect of UCS intensity. According to the F values there were significant differences in the latency and percentage of responses in acquisition and no significant differences for either response measure in extinction. On the other hand, the analyses of response angle failed to reveal significant differences in acquisition and extinction. Individual t test comparisons of the arc-sine transform of percentage acquisition scores yielded a P value of less than .01 for the NV50 v. NV200 comparison and failed to reach the .05 level for all other intergroup comparisons. This indicates that the significant F value obtained for UCS intensity is primarily determined by the superior performance of Group NV200 to Group NV50. Since the mean level of performance of the V groups was close to the ceiling of response measurement, it is not surprising to observe that the 5% increase in the mean percentage of responses from Group V50 to Groups V100 and V200 was not statistically significant. In the analyses of the log transform of latency acquisition scores, t test comparisons yielded Ps of less than .05 for four of the six intergroup comparisons (viz., V50 vs. V100, $P < .02$; V50 vs. V200, $P < .001$; NV50 vs. NV200, $P < .05$; NV100 vs. NV200, $P < .01$). The failure of the NV50 vs. NV100 t test comparison to reach statistical significance indicates that the observed increase in response latency from Group NV50 to Group NV100 was not a significant inversion in the observed UCS intensity-response latency function. Row 2 indicates that voluntary blink instructed *S*s had significantly shorter mean response latencies, higher levels of percentage of responses, and sharper lid closures than nonvoluntary blink instructed *S*s in both acquisition and extinction. In Row 3 it is shown that the source of variation attributable to differences in the two recording systems failed to reach significance for any of the response measures in both acquisition and extinction. Row 4 presents the interaction effects of UCS intensity and instructional set. The failure of the F values to reach statistical significance in acquisition or extinction suggests that variations in UCS intensity had

TABLE 2

ANALYSES OF VARIANCE OF LATENCY, PERCENTAGE, AND ANGLE OF EYELID RESPONSES DURING ACQUISITION AND EXTINCTION

Source of Variation	*df*	Latency		Percentage		Angle	
		Acquisition	Extinction	Acquisition	Extinction	Acquisition	Extinction
UCS intensity (I)	2	9.85***	2.93	4.29*	1.26	.47	1.48
Instructional set (S)	1	94.32***	58.26***	82.98***	4.94*	26.12***	6.85*
Recording system (R)	1	.06	.11	.53	.10	1.32	1.97
I × S	2	2.77	2.85	1.55	.45	.70	.77
I × R	2	.65	.82	.54	1.19	.48	.46
S × R	1	1.94	.61	.07	.08	6.67*	4.92*
I × S × R	2	1.70	3.70*	.06	1.94	.16	.17
Error	132	(.0034)	(.0061)	(112.17)	(443.52)	(.0405)	(.0598)

Note.—Number in parentheses are error mean squares. The latency and angle analyses were performed on the log transforms of individual mean latencies and individual mean response angles; the frequency analyses were performed on the arc-sine transforms of individual percentage scores.

* $P < .05$.
** $P < .01$.
*** $P < .001$.

similar effects under both instructional sets. Examination of the remaining *F* values in the table indicates that significant *F*s were obtained in Rows 6 and 7. Row 6 shows there were significant interaction effects of instructional set and recording system on the form of the response in acquisition and extinction. These interaction effects arose from the fact that in both acquisition and extinction the mean response angle under V instructions was greater in Recording System I than in Recording System II, whereas under NV instructions the mean response angle was greater under Recording System II than under Recording System I. The significant *F* value in Row 7 indicates that the second-order interaction of UCS intensity, instructional set, and recording system had a significant effect on response latency in extinction. However, the significant *F* values obtained in Rows 6 and 7 appear to be devoid of psychological significance.

Discussion

The principal findings of this investigation were: (*a*) *S*s under voluntary instructional set exhibited responses of shorter latency, attained higher levels of percentage of responses, and had sharper lid closures than *S*s under nonvoluntary instructional set in both acquisition and extinction; (*b*) as UCS intensity increased it was observed that in acquisition response latency decreased, percentage of responses increased, and sharpness of lid closure showed no systematic change; and (*c*) there were no significant interaction effects between instructional set and UCS intensity on the latency, percentage, or form of the response in acquisition or extinction.

The first finding indicates that the response characteristics of voluntary blink instructed *S*s are markedly different from those of *S*s conditioned under conventional neutral instructions. It also indicates, so far as acquisition is concerned, considerable similarities between the response characteristics of voluntary blink instructed *S*s and those attributed to judged voluntary responders by Spence and his associates (Spence, 1953; Spence & Ross, 1957, 1959; Spence & Taylor, 1951). The latency distributions of voluntary blink instructed *S*s further confirms the correspondence to judged voluntary responders reported by Hartman, Grant, and Ross (1960) and are also strikingly similar to judged voluntary responses without a ready signal, reported by Hartman and Ross (1961). The observation that the response characteristics of voluntary blink instructed *S*s were maintained in extinction suggests that their responses became conditioned under the acquisition procedure. (An alternative possibility is that voluntary blink instructed *S*s interpreted the instructions to require that they blink to the CS even in the absence of the UCS.) As yet, conditioning studies which have recovered judged voluntary responders from their data have not attempted to investigate the behavior of these *S*s in extinction. If however, their extinction curves showed a greater resistance to extinction, it would suggest that the term "voluntary responder" is very inappropriate and that "highly conditionable" *S*s (i.e., *S*s with higher learning rate parameters) would be a better one.[2]

The second finding, that increasing UCS intensity results in increasing percentage of responses in acquisition, is a conventional result. However, the fact that increasing UCS intensity produces systematic decreases in response latency

[2] To explore such a possibility, the latency criterion suggested by Spence and Ross (1959) was applied to *S*s in the NV groups (i.e., *S*s who gave 50% or more of their CRs before 300 msec. were classified as "voluntary" responders). Since a ready signal was not employed in the present investigation this latency criterion would presumably be a conservative criterion for recovering voluntary responders (Hartman et al., 1960). Nevertheless, 3 voluntary *S*s were recovered from Group NV50, 2 from NV100, and 3 from NV200. In acquisition, the mean percentage of CRs for these *S*s were 84.8% for NV50, 86.4% for NV100, 90.8% for NV200, and in extinction the means were 66.7, 47.5, and 49.0%, respectively. Though these observations were based on a small and unequal number of *S*s, the mean percentage of CRs of these voluntary *S*s was considerably greater in both acquisition and extinction than that obtained under each of the NV conditions with unselected *S*s.

has methodological implications for eyelid conditioning studies which have varied UCS intensity and concurrently employed a latency criterion to reject *Ss* as voluntary responders. As UCS intensity increases such a procedure would classify increasing numbers of *Ss* as voluntary responders. (The possibility also exists that the observed latency-UCS intensity relationship would be different for studies employing a ready signal.) It could be maintained, of course, that increases in UCS intensity simply increases the incidence of voluntary responders. Spence and Ross (1959) implied this argument when they stated that ". . . a relatively strong air puff was employed in order to elicit a large number of voluntary type responses" (p. 378). Such reasoning brings into sharp focus another difficulty in current attempts to classify *Ss* as voluntary responders on the basis of the topography of their responses. Unless it has been empirically determined that the response characteristics attributed to voluntary responders, and employed as criteria in their selection, are *not* a function of independent variables employed in the conditioning procedure, the equally valid assumption can be made that *Ss* will be rejected as voluntary responders for behavior lawfully related to different independent variables and/or different values of the independent variables. Spence and his associates may choose to delete these *Ss* on the basis of response latency in order to simplify their descriptive problem within the Hull-Spence theoretical framework. However, the basic difficulty is that one is forced into the logically difficult position of having to employ criteria in the selection of these *Ss* which have been shown to be invariant, not only for a restricted set of conditioning parameters, but for the infinite population of parameters. In the absence of such cross validation of these criteria, one is confronted with the possibility of erroneously classifying (or failing to classify) different numbers of *Ss* as voluntary responders for behavior functionally related to the conditioning parameters employed. The failure of response latency to satisfy this condition of invariance is indicated by the presently observed latency-UCS intensity relationship and further, by the findings of Hartman, Grant, and Ross (1960) that the response latency of voluntary blink instructed *Ss* remained invariant for massed and distributed conditioning trials with a ready signal, but that a shorter response latency was observed in massed than spaced conditioning trials in the absence of a ready signal. Whether objective measures of sharpness of lid closure will meet with similar difficulties remains an empirical question.

The third finding of failing to obtain a significant interaction effect of instructional set and UCS intensity on percentage of responses is related to the observation of Spence and Ross (1959) that judged voluntary responders follow different laws of conditioning than nonvoluntary responders. Specifically, it has been maintained that the percentage of responses of these *Ss* do not vary with changes in UCS intensity. However, if judged voluntary *Ss* are simply *Ss* with higher conditioning parameters and they are selected from experiments in which there were differences in UCS intensity, then one might expect to observe no differences since they are performing near the ceiling of response measurement. The fact that the response angles of voluntary blink instructed *Ss* followed similar laws for variation in UCS intensity as nonvoluntary instructed *Ss* suggests that a specified criterion of response angle could be employed to discriminate the two instructional groups independent of UCS intensity. The observed invariance requires proof of the null hypothesis. Although this cannot be done, it is proper to indicate that with 144 *Ss* the power of the experiment was considerable, so that if the latency, frequency, and form of responses under the two instructional sets did follow different laws for UCS intensity the differences must not have been appreciable.

Summary

The effects of voluntary (V) and nonvoluntary (NV) blink instructions on the latency, frequency, and form of eyelid responses were compared under classical conditioning and extinction procedures, with UCS intensity as a parameter. It was found that: (*a*) The *Ss* under V instructions gave responses of shorter latency, attained higher percentages of responses, and had sharper lid closures than *Ss* under NV instructions in both acquisition and extinction. (*b*) As UCS intensity increased, acquisition response latency decreased, percentage of responses increased, and the form of lid closure failed to show systematic change. (*c*) There were no significant interaction effects between instructional set and UCS intensity.

It was concluded that: (*a*) the responses of the V *Ss* became conditioned under the acquisition procedure; (*b*) studies that have varied UCS intensity and concurrently employed a latency criterion to reject *Ss* as voluntary responders may have introduced a methodological error; (*c*) V *Ss* may not follow appreciably different laws to variations in UCS intensity than NV *Ss*.

References

Bartlett, M. S. The use of transformations. *Biometrics*, 1947, **3,** 39–52.

Grant, D. A. Sensitization and association in eyelid conditioning. *J. exp. Psychol.*, 1943, **32,** 201–212.

Grant, D. A., & Norris, E. B. Dark adaptation as a factor in the sensitization of the Beta response of the eyelid response to light. *J. exp. Psychol.*, 1946, **36,** 390–397.

Hartman, T. F., Grant, D. A., & Ross, L. E. An investigation of the latency of "instructed voluntary" eyelid responses. *Psychol. Rep.*, 1960, **7,** 305–311.

Hartman, T. F., & Ross, L. E. An alternative criterion for the elimination of "voluntary" responses in eyelid conditioning. *J. exp. Psychol.*, 1961, **61,** 334–338.

Moore, J. W., & Gormezano, I. Yoked comparisons of instrumental and classical eyelid conditioning. *J. exp. Psychol.*, 1961, **62,** 552–559.

Spence, K. W. Learning and performance in eyelid conditioning as a function of the intensity of the UCS. *J. exp. Psychol.*, 1953, **45,** 57–63.

Spence, K. W., & Ross, L. E. Experimental evidence on the relation between performance level in eyelid conditioning and anxiety (drive) level. *Off. Naval Res. tech. Rep.*, 1957, No. 5. (Contract No. N9-ONR-93802).

Spence K. W., & Ross, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58,** 376–381.

Spence, K. W., & Taylor, J. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42,** 183–188.

(Received March 22, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 495–498

TENDENCIES TO EYE MOVEMENT AND PERCEPTUAL ACCURACY[1]

HERBERT F. CROVITZ AND WALTER DAVES

Veterans Administration Hospital, Durham, N. C.

Duke University

The present study deals with one possible role of sensory-motor factors in perception. While the movement of the eyes leads to new visual stimulation which may be relevant in perceiving, the neural activity responsible for an eye movement may itself play a role in perceptual integration whether or not an eye movement occurs and whether or not the sensory feedback effects of an eye movement are supplied to the locus of perceptual integration. This neural state tending to produce an eye movement and having a possible role in perceptual processing will be called a "tendency to eye movement."

The postulated process may be related to the process underlying the use of the term tendencies to eye movement as used previously in psychology. For example, Woodworth (1938) cites a theory relating eye movements to the Muller-Lyer illusion which ". . . admits that actual movements do not occur in all cases, but assumes that a tendency to such movement is sufficient to give the impression of length" (p. 645). Recently the term has been used in perceptual theory by Gaffron (1950), Heron (1957), and Crovitz (1960).

In normal visual experience, perception may be affected by both the feedback supplied by eye movement and by the neural state which initiated the movement. Using a tachistoscope, however, when exposure time is shorter than the latency of eye movement, the eye movements which occur lead to the fixation of the blank postexposure field. The *first* eye movement which occurs must be the result of the tendency to eye movement existing at the initiation of the movement. If, as assumed, there is a relationship between the tendency to eye movement qua neural state, and perception, a relationship should exist between some property of the initial eye movement, as an index of the tendency, and some property of perception. It would appear reasonable that a property in which a relationship could be shown would be the directionality of eye movement and perceptual accuracy; i.e., a congruence might be expected between the direction of the initial eye movement and the direction of the more accurate field.

METHOD

Stimulus materials.—The stimulus material[2] consisted of three sets of eight cards, each card with numerals appearing at 3°, 5°, and 7° of visual angle of both the left and right stimulus fields of a Gerbrands mirror tachistoscope. The tachistoscope was fitted with an

[1] The authors are indebted to Karl Zener for his valued advice in the formulation and execution of this study, to Robert L. Green who made available part of the eye movement recording apparatus used, and to Paul G. Daston for critical reading of the report.

[2] The perceptual situation chosen was determined by requirements of a replication of an unpublished study on perceptual anisotropies and laterality done by Hilborn at Duke University. The present report, however, does not relate to the specifics of Hilborn's problem.

TABLE 1

FREQUENCY OF OCCURRENCE OF EACH COMBINATION OF ACCURACY DIFFERENCE AND EYE MOVEMENT DIRECTION FOR EACH VIEWING CONDITION

Viewing Condition	More Accurate Field	Eye Movement Direction			
		Rt.	Lt.	No.	Total
Both eyes	Right	30	3	10	43
	Left	2	32	7	41
	Equal	5	14	9	28
	Total	37	49	26	
Right eye	Right	39	4	12	55
	Left	6	8	13	27
	Equal	8	11	11	30
	Total	53	23	36	
Left eye	Right	18	8	11	37
	Left	4	26	13	43
	Equal	3	17	12	32
	Total	25	51	36	

Note.—Rt. = Right; Lt. = Left; No. = None.

eye aperture occluder such that the line of numerals across the whole field could be viewed binocularly or with the left or the right eye alone. Each set of eight cards used in the three viewing conditions (both eyes, left eye alone, right eye alone) had each of the numerals (2–9) in each of the six positions once. Otherwise the positions for the numerals were randomly selected. The numerals, executed in India ink using a Zephyr lettering kit, stood about $\frac{3}{16}$ in. high, and subtended a visual angle of about 30′.

Procedure.—Cards were presented in a fixed random order, binocular and monocular trials intermixed. On each trial *S* adjusted the eye aperture occluder according to *E*'s instructions, viewing with both eyes, left eye alone, and right eye alone for a total of eight trials each. A constant exposure time of 100 msec. was used, and since the latency of voluntary eye movements is somewhat longer than this (Woodworth, 1938) such postexposure eye movements could not lead to fixation of a numeral.

Recording technique.—An electro-oculograph technique was used to record initial postexposure eye movements. Eye movement potentials were amplified by a Grass Model III-D high efficiency EEG machine equipped with converter-demodulators (Grass Model CD-3) in order to provide dc voltage input to the pen. Electrodes were made of jeweler's "high fine" silver disks, 1 mm. thick by 8 mm. diameter, plated with silver chloride and encased in silver cups, in general agreement with the method suggested by Ford and Leonard (1958).

Electrodes were placed bitemporally for recording the horizontal components of the eye movements. A second channel was connected to the tachistoscope timer such that a sharp spike was produced on the moving paper record each time the stimulus field was flashed on and each time it was flashed off, allowing for and leading to a check that the initial eye movement recorded began after the stimulus material flashed off. Paper speed was 30 mm/sec.

Before the electrodes were taped into position, the skin sites were washed with acetone and lightly sanded with fine grade sandpaper in order to reduce the resistance between the pair of electrodes. While the resistance was not measured directly (an extremely low voltage ohmmeter was not available and the phosphene effect of passing a current through the electrodes might have produced undue anxiety in the naive *S*s) the resistance was considered sufficiently low when a calibration eye movement of 7° produced a pen deflection of at least 5 mm. on the eye movement channel with the gain setting used.

Instructions.—The instructions to *S* were to fixate a centrally placed dot in the fixation field upon a "ready" signal, to refrain from moving his eyes until the stimulus appeared, to move his eyes to whatever numeral or numerals he pleased when the stimulus appeared.

The *S* was told to write on an answer sheet after each trial the numerals he had seen in the positions in which they had appeared.

Subjects.—Fourteen naive college students served as *S*s.

RESULTS

The data support the hypothesis that a congruence in direction exists between the initial postexposure eye movement and the more accurate field on trials in which there was both an eye movement and an accuracy difference between the fields. There were more congruent than noncongruent trials ($t = 5.05$, $P < .01$), and this relationship held for 13 of the

14 *S*s (a binomial test giving $P < .01$).

This congruence also appears in each viewing condition individually. Table 1 presents the frequency of occurrence of trials with each combination of eye movement direction and accuracy difference for each viewing condition.

Disregarding perceptual accuracy, in both monocular sets of trials there are twice as many initial eye movements in the direction of the viewing eye than in the opposite direction; i.e., with right eye alone there are more initial eye movements to the right than to the left ($t = 4.12$; $P < .01$) and with left eye alone there are more initial eye movements to the left than to the right ($t = 5.36$; $P < .01$).

Scorable initial postexposure eye movements (in a latency range of from 150 msec. to 1000 msec. after stimulus onset) did not occur on about 30% of the trials. This can, in part, be attributed to the electro-oculograph method used. It was found to be subject to interference from eye blinks and occasional galvanic skin response potentials when such occurred in the latency range chosen.

Perceptual accuracy (defined as correct numeral reported in its correct location) was moderately high, with 52.9% of all reported numerals correct. However, *S*s recorded only 35.8% of all reportable numerals.

Discussion

The results of the present experiment reveal a relation between relative accuracy of perception of numerals on either side of the fixation point and the direction of the initial postexposure eye movement. The conditions of the experiment effectively precluded any visual feedback from eye movement which could lead to differential accuracy on the two sides of the field.

The significance of the present finding lies in the clear indication that other processes than visual feedback are significantly related to accuracy of report of visual material. However, the design of the study does not permit differentiation between the effects of tendencies to eye movements and the effects of possible kinesthetic feedback from the initial eye movement, a task for later empirical determination. Further, one explanation of the findings of a visual nature might lie in the possibility that the clearer perception of numerals on one side of the field, for whatever reason, might lead to an actual eye movement in that direction of relatively short latency. Congruence occured on 85% of the trials in which the latency of initial postexposure eye movements was from 150 to 175 msec. Whether such short latencies are consistent with an alternative explanation in terms of an underlying visual discrimination process is a question for more discriminating empirical determination.

The clear demonstration of an association between accuracy of report and eye movement direction urgently raises the need for clarification of the detailed ways in which tendencies to movement or other relevant variables may affect perception and calls for tests of alternative hypotheses. For instance, one problem is whether the motor tendency directly affects properties of the visual reaction itself or is more intimately related to the process of accurate report.

A secondary positive finding which also invites specific explanation is that initial postexposure eye movements tend to be made to the left when viewing with the left eye and to the right when viewing with the right eye. At a physiological level, the movement of the viewing eye is related to action of the external rectus muscle, and movements are made toward stimuli exciting nasal retina; while at a psychological level a possible relevant fact might be that *S* adjusted the eye aperture occluder and was "aware" that he viewed with a given

eye alone, eye movements being in the direction away from the occluded aperture. Again an experimental determination between alternative explanations might give insight into the details of the processes involved.

Regardless of ultimate explanation, the findings of the present experiment indicate the necessity of taking into account direction of initial postexposure eye movements as related to accuracy differences between the fields in tachistoscopic studies, and gives evidence that the presentation of stimuli at exposure times short enough to rule out fixation eye movements does not eliminate the possible effects of eye movements, or tendencies to eye movements, upon the subsequent report.

Summary

The direction of initial postexposure eye movements was studied in a tachistoscopic situation in which a row of numerals appeared across the visual field and no eye movement occurred until the cessation of stimulation. A congruence was found between the direction of the initial eye movement and the side of the visual field more accurately perceived. This finding supports the hypothesis that differential tendencies to eye movement are associated with differential accuracy. A secondary finding was that, in monocular viewing, there were more eye movements to the side of the viewing eye.

References

Crovitz, H. F. Patterns of relative localiza tion of an odd element within a visua-grouping as a function of laterality charl acteristics and tendencies to eye movement-Unpublished doctoral dissertation, Duke. University, 1960.

Ford, A., & Leonard, J. L. Techniques for recording surface biolectric direct currents. *USN Electron. Lab. Rep.*, 1958, No. 839.

Gaffron, M. Right and left in pictures. *Art Quart.*, 1950, **13,** 312–331.

Heron, W. Perception as a function of retinal locus and attention. *Amer. J. Psychol.*, 1957, **70,** 38–48.

Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received March 27, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 499–504

THE EFFECT OF NONSENSE-SYLLABLE COMPOUND STIMULI ON LATENCY IN A VERBAL PAIRED-ASSOCIATE TASK[1]

BARBARA S. MUSGRAVE[2]

University of Massachusetts

The present experiment was concerned with effects of convergent, associated-nonassociated, and divergent stimulus compounds on latency of responses in a verbal paired-associate (PA) task. The paradigm for convergent compounds is to train *S*s to respond with the same response to each of the stimulus elements presented separately, and then to present the elements in combination. For divergent compounds, the procedure is the traditional conflict paradigm in which *S*s are trained to respond with different responses to each of the stimulus elements, which are then presented together. The paradigm for associated-nonassociated compounds is to train *S*s to respond to a stimulus element and then to present that element accompanied by a new, previously unpresented, element.

According to a fuller review presented elsewhere (Musgrave, 1960), findings with stimulus compounds in relatively simple learning tasks suggest that convergent compounds might facilitate performance, such facilitation being explained in terms of summation of response strengths. On the other hand, findings of inhibitory effects of response conflict suggest that divergent compounds might hinder performance. Evidence concerning the effects of including a novel element in the stimulus are somewhat ambiguous but may be interpreted that performance with associated-nonassociated compounds would be intermediate between that for convergent and that for divergent compounds.

Only Shepard and Fogelsonger (1913), using a PA technique, have investigated the effects on latency and recall of combining into verbal stimulus compounds the single stimuli of experimentally established verbal associations. Unfortunately, none of their experiments was free of one or more methodological shortcomings. The present study, therefore, was designed to refine and extend information on effects of convergent, associated-nonassociated, and divergent compounds in a verbal PA task.

METHOD

Design.—In addition to the experimental conditions using the three different types of compounds, three control conditions were used: Continued Singles, in which a single syllable of the training phase was the stimulus of the test phase; New Singles and New Doubles in which the stimuli of the test phase had not appeared during the training phase. The Continued Singles condition permitted comparisons between compounds and elements composing the compounds. The New Singles and New Doubles were controls for effects of such factors as familiarization with the mode of presentation of the stimuli and experience with the PA technique.

[1] This paper is based upon a dissertation submitted to the Department of Psychology of the University of Massachusetts in partial fulfillment of the requirements of the PhD degree, and was supported by a research fellowship, MF-6860-C, from the National Institute of Mental Health, United States Public Health Service. The writer is greatly indebted to A. E. Goss under whose direction this investigation was conducted. Thanks are due also to J. L. Myers for help with the statistical design.

[2] Now at Smith College.

Because written or spoken words are typically sequential, order of the stimulus elements in the sequence may be an important variable, with the element nearest the response the more effective (Howes & Osgood, 1954; Musgrave, 1958). Therefore, order was introduced as a second variable and the nonsense-syllable elements of each of the kinds of compound appeared in both orders. Order of the syllables, when each is the stimulus for the same response, as in the Convergent condition, involves nothing more than counterbalancing for occurrence in the first or second position of the particular syllables used. In both associated-nonassociated and divergent compounds, however, training differences are involved. Consequently, order is not orthogonal to type of compound. For convenience, the two orders used were labeled I and II. Order I refers to compounds in which the syllable which appears first, reading from left to right, is the stimulus whose response was to be considered correct on the test trials. Order II refers to compounds in which this critical stimulus occurs second.

Subjects.—Seventy-two undergraduate women were tested individually, being assigned as they appeared to two groups of 8 *Ss* each to serve in the two counterbalanced Convergent conditions, to two groups of 4 *Ss* each to serve in the two counterbalanced New Doubles conditions, and to six groups of 8 *Ss* each to serve in the remaining six conditions. Forty-eight additional *Ss* were tested but discarded because of failure to meet one or more of three criteria of performance on the single-syllable stimuli of the training lists. These criteria are described in the section on procedure.

Apparatus.—The paired associates were presented by means of a modified Hull memory drum. With the opening of a left-hand shutter, the stimulus member of each pair was presented alone for 2 sec., then in addition to the stimulus, with the opening of a right-hand shutter, the response member was presented for 2 sec. The interpair interval was 4 sec. Latency of responses was obtained by means of a Gerbrand's electronic voice key and a Hunter Klockounter. The sound of the first letter of the spelled response, acting through the voice key, stopped the Klockounter.

Paired associates.—Forty nonsense syllables, selected from Glaze's (1928) list of syllables with high association values (100% and 93%) and of low similarity to each other, were used. In order that test materials for all conditions would contain the same syllables, differing training lists were used in different conditions.

The training materials included three kinds of items: (*a*) the basic list of single-syllable stimulus and response pairs, (*b*) pairs of associated-nonassociated stimuli and single-syllable responses, and (*c*) either single-syllable or associated-nonassociated stimuli alone, without their appropriate responses. The basic list consisted of four pairs made from four single-syllable stimuli and two single-syllable responses. In the associated-nonassociated pairs, the four stimuli in the basic list were each accompanied by a second syllable which had not previously appeared and did not appear again. These associated-nonassociated pairs were used to prevent formation of a set-for-single-stimuli by providing experience with compound stimuli (Andreas, 1958). Also, they served to accustom *S* to respond to stimulus compounds and to respond with only one or the other of the two responses of the learning task. The third type of item—the stimulus presented alone—was used to accustom *S* to a response period of 4 sec. rather than 2 sec. This lengthened response period seemed desirable to use on the first test trial following training because preliminary studies had shown that 2 sec. was insufficient time for anticipatory responses under Divergent, New Doubles, and New Singles conditions.

The training materials consisted of 116 items, divided into four lists of 24 items each, and a fifth list which began with 20 training items and ended with the items for the four test trials. Before each *S* arrived all five lists for the appropriate condition were arrayed on the drum. After the first list of 24 was presented, the drum was moved along its axle and rotated to the starting position for presentation of the next list of 24 and so on.

The items for the four test trials consisted of a single stimulus-response pair presented four successive times. The stimuli for the New Singles and Continued Singles were single syllables. For the remaining six conditions, the stimuli were two-syllable compounds. On Trial 1, the stimulus was followed by a blank. On the remaining three trials, the stimulus was followed by responses designated correct by *E*.

Procedure.—The order in which the conditions were presented was used to assign *Ss* to groups. The order was random with the restriction that every condition should be presented once before any condition was presented twice and so on. Within this restriction, the 54 different counterbalanced combinations of training and test materials (72 combinations in all) were assigned to a

schedule sheet with the aid of a table of random numbers.

Each *S* was seated in front of the apparatus, asked to hold the microphone close to her mouth and to speak the word "Hello" when each of the shutters opened to reveal the stimulus and then the response. This procedure was explained to *S* as necessary for adjusting the apparatus for her individual voice, and permitted coaching her in speaking loudly and clearly and in responding twice on every trial, once before and once after the opening of the second, or response, shutter. The *E* then sat out of sight behind the apparatus and read the instructions. In addition to general information concerning the PA task, instructions included the specific information that there would be only two response syllables. What these syllables would be was told to *S* six times. The *S* then practiced with the training materials. The final list, as mentioned above, ended with four successive presentations of the particular compound stimulus or, in the case of the New Singles and Continued Singles conditions, the single-syllable stimulus which constituted the test stimulus. No new instructions or other comments preceded the presentation of the test trials. On the first occurrence of the test stimulus, no response member was presented. Thus, on this trial, as the test stimulus was presented alone for 4 sec., it was possible to measure latencies up to 4 sec. rather than up to 2 sec. On Trials 2, 3, and 4 of the test stimulus, the correct response member was exposed for 2 sec. Thus, Test Trials 1 and 2 were, in effect, recall trials, while Trials 3 and 4, following the initial presentation of the correct response on Trial 2, were learning trials.

Following completion of each *S*'s participation, her performance was evaluated on the basis of three criteria: (*a*) correct responses on the last two presentations of each of the critical training syllables, i.e., the syllables which in the Convergent and Divergent conditions were paired together as compound stimuli in the test trials, or the corresponding stimuli in the training lists of the other conditions; (*b*) response latencies within .300 sec. of each other on the last occurrence of the two critical syllables; (*c*) an equal number of correct responses, plus or minus one response, for the 11 times each syllable appeared in the last three lists. The first of these criteria was included to insure that *S*s had actually learned the correct responses to the critical syllables before entering the test phase.

The second and third criteria were of particular importance in the Divergent (conflict) condition, as they were designed to insure that responses to the two syllables were at equal or near equal strength, thus guarding against the possibility that choice of responses reflected nothing more than differential strengths. To equate selection of *S*s through conditions, these two criteria were applied not only to the Divergent but to all other conditions as well. In connection with the third criterion, it was assumed from preliminary studies that the numbers of correct responses would be fairly stable across conditions. However, it was decided that any numbers of correct responses which deviated from the numbers of correct responses for the majority of *S*s, even if equal between the two critical syllables, would be sufficient cause for replacing an *S*. If the performance of any *S* did not meet these criteria, her protocol was discarded and the next *S* to appear was assigned to the same condition.

Results

Training Trials

The single syllables which were to appear in the test compounds appeared in the last three training trials 11 times. The *S*s of all groups responded correctly to these syllables either 10 or 11 times.

Latencies on the last training trial, the criterion trial, for each of the two single syllables which were to be combined in the test compounds were averaged. Means and *SD*s of these averaged latencies for all eight conditions are presented in Table 1. The latencies of both criterion and test trials exhibited marked heterogeneity of variance. Accordingly, for the analysis of variance, these data were transformed into their reciprocals (speed of response evocation). The analysis of variance of the criterion-trial data indicated that there were no significant differences among conditions at the end of the training trials ($F < 1$; $df = 7/64$).

TABLE 1

MEANS AND *SD*s OF LATENCIES ON THE CRITERION TRAINING TRIAL AND FOUR TEST TRIALS FOR ALL CONDITIONS

Condition	Training		Test							
	Criterion Trial		Trial 1		Trial 2		Trial 3		Trial 4	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Continued Singles (Cont. S.)	0.958	.090	0.902	.123	0.960	.212	0.904	.114	0.917	.145
Convergent (Conv.)	0.906	.100	1.113	.205	1.064	.212	1.000	.179	1.001	.263
Asso.-Nonasso. I (A-NA I)	0.911	.071	1.196	.237	1.099	.249	1.028	.173	0.980	.167
Asso.-Nonasso. II (A-NA II)	0.961	.179	1.233	.279	1.088	.167	1.006	.164	1.065	.219
Divergent I (Div. I)	0.981	.131	1.236	.164	1.635	.700	1.267	.363	1.121	.198
Divergent II (Div. II)	0.941	.188	1.240	.501	1.162	.255	1.451	.625	1.129	.278
New Singles (New S.)	0.928	.084	1.969	.852	1.620	.779	1.134	.400	0.990	.155
New Doubles (New D.)	1.006	.095	2.680	1.099	1.812	.938	1.162	.819	1.082	.739

Test Trials

Latencies on all four test trials.—Means and *SD*s of latencies for all eight conditions for each of the four test trials are presented in Table 1. The *F*s for Conditions, Trials, and Trials × Conditions were significant at less than .001 (Table 2). Because of the significant Trials × Conditions interaction, separate analyses of variance were carried out for each trial. Differences among Conditions were significant for Trials 1 ($P < .001$), 2 ($P < .001$), and 3 ($P < .01$) but were not significant for Trial 4.

TABLE 2

SUMMARY OF DUNCAN RANGE TESTS

Cond.	Cont. S.	Conv.	Div. II	A-NA I	A-NA II	Div. I
Trial 1						
Conv.	.05	—	—	—	—	—
Div. II	—	—	—	—	—	—
A-NA I	.05	—	—	—	—	—
A-NA II	.05	—	—	—	—	—
Div. I	.05	—	—	—	—	—
New S.	.01	.01	.01	.05	.05	.05
New D.	.01	.01	.01	.01	.01	.01
Trial 2						
New S.	.01	.01		.05	.05	
Div. I	.01	.01		.05	.05	
New D.	.01	.01		.01	.05	
Trial 3						
New D.	.05	—		—	—	
Div. I	.01	.05		—	—	
Div. II	.01	.01		.05	.05	

Note.—Conditions in the margins both from top to bottom and left to right of each table are listed in the order of magnitude of means for each trial, except for a reverse positioning of A-NA II and A-NA I in the table for Trial 3. A-NA II and A-NA are not reversed on Trial 2 due to a shift effected in the direction of the means by the transformation from latencies to speeds.

The results of a Duncan multiple range test of the significances of differences between all pairs of means of speeds (reciprocals of latencies) on each of the four trials are summarized in Table 2.

An analysis of variance for each condition separately assessed the effects of order and of trials both for the Divergent and Associated-Nonassociated conditions in which order of syllables was based on training differences and for the Convergent and New Doubles conditions in which order reflected only a counterbalancing of particular syllables. In no case was order a significant variable ($F < 1$ in all cases). However, across trials, significant decreases in latencies occurred under Associated-Nonasso-

ciated, New Doubles, and New Singles conditions.

Number correct on recall trials.—For the Convergent, Associated-Nonassociated, and Continued Singles conditions, which had been trained with only one response syllable designated as correct, all *S*s responded correctly on Test Trials 1 and 2. For the Divergent, New Doubles, and New Singles conditions *either* of the two trained response syllables was scored as correct on the recall trials. On Trial 2 1 *S* in Divergent 1 responded after the opening of the second shutter had revealed the correct response and, hence, her response was marked incorrect. In the New Doubles condition, 1 *S* failed to respond on Trial 1, and a second *S* made no response on either recall trial. One *S* in the New Singles condition made no response on Trial 1 and waited to read the response syllable on Trial 2; these responses were scored incorrect.

Number correct on learning trials.—After exposure of the stimulus on Trial 2, the correct response syllable had been exposed. Thereafter, on Trials 3 and 4, only this syllable was scored as correct. Only *S*s in the Divergent condition failed to respond correctly. Three *S*s responded with the incorrect syllable on Trial 3, 1 responded with the incorrect syllable on both Trials 3 and 4, and another failed to respond on Trial 3 before the second shutter opened.

Discussion

The expected rank order of increasing latencies for the experimental conditions of Convergent, Associated-Nonassociated, and Divergent was confirmed. However, the Convergent condition did not facilitate performance relative to the Continued Singles. That facilitation was possible is indicated by the fact that latencies for the Continued Singles condition were shorter on all four test trials than those for any condition on the last training trial. In contrast, latencies for the Convergent condition were longer than those for any condition on the last training trial, and were longer than those for the Continued Singles condition throughout the four test trials. Either the convergent associations did not summate, or the effects of summation were outweighed by factors impairing performance. For example, it may have taken longer to look at and perhaps to read two-syllable stimuli. Or the poorer performance may have been due to dissimilarity between presentation of one nonsense syllable alone and presentation of that syllable simultaneously with another nonsense syllable. The *S*s had had experience with each of the test syllables presented with another syllable but their training had not included presentation of the test syllable with another syllable to which a response had been associated experimentally.

The Associated-Nonassociated conditions resulted in consistently poorer performance than the Convergent condition, but the differences were slighter than expected. Possibly presentation of Associated-Nonassociated compounds during training had accustomed *S*s to quick selection of the associated syllable and immediate disregard of the nonassociated syllable. In the introduction of a novel stimulus element, there are two sources of novelty. The introduction may be novel, the element may be novel, or both. With the procedure followed here, the introduction of a novel element was no longer completely novel, nor was doubleness as such. However, the nonassociated element of the first test stimulus was novel in that it had not been presented previously in the experimental situation. Presumably, then, the novelty of this element was chiefly responsible for the differences between the Associated-Nonassociated conditions and the Continued Singles and Convergent conditions.

Latencies on Trial 1 did not fall as low for the New Singles condition as for the New Doubles, and the slope of the

improvement curve was steeper for the New Singles. These results indicate the desirability of enlarging available information concerning the difficulty of multi-element stimuli, possible interaction of number of stimulus elements with length of list, and so on.

Results for the Divergent conditions differed sharply and in several ways from those for the other five conditions. In general, what is thought of as typical conflict behavior was exhibited. In addition to long latencies, responses of *S*s in the Divergent condition showed vacillation, providing instances not only of response blends but also of response alternation from trial to trial even after the correct response had been presented.

Summary

The present study was concerned with the effects on response latency of convergent, divergent, and associated-nonassociated nonsense-syllable compounds. Order of the syllables in the divergent and associated-nonassociated compounds was a second, nonorthogonal variable. Continued Singles, New Singles, and New Doubles conditions served as controls.

On the first of four test trials, all experimental compound conditions were slower than Continued Singles and faster than the New Singles and the New Doubles conditions. Although differences among the experimental compound conditions were slight, the rank order of increasing latencies of the Convergent, Associated-Nonassociated, and Divergent conditions was that which had been predicted.

On the remaining three test trials, Continued Singles did not improve, suggesting that asymptotic latencies had been reached. In contrast, latencies for the Convergent condition and for the two Associated-Nonassociated conditions decreased moderately, while those for the New Singles and New Doubles conditions decreased markedly. Only the two Divergent conditions did not exhibit either asymptotic latencies or reasonably steady improvement.

References

Andreas, B. G. Motor conflict as a function of motivation and amount of training. *J. exp. Psychol.*, 1958, **55**, 173–178.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

Howes, D., & Osgood, C. E. On the combination of associative probabilities in linguistic contexts. *Amer. J. Psychol.*, 1954, **67**, 241–258.

Musgrave, B. S. Context effects on word associations using one-word, two-word, and three-word stimuli. Paper read at Eastern Psychological Association, Philadelphia, April 1958.

Musgrave, B. S. Response speed, recall, and learning with nonsense-syllable compound stimuli. Unpublished doctoral dissertation, University of Massachusetts, 1960.

Shepard, J. F., & Fogelsonger, H. M. Studies in association and inhibition. *Psychol. Rev.*, 1913, **20**, 290–311.

(Received March 27, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 505–509

COMBINATION OF DRIVE AND DRUG EFFECTS

JOSEPH MENDELSON AND DALBIR BINDRA

McGill University

Does the effect of a drug on a trained response depend upon the current drive level of *S*? In other words, is the performance of a given response more vulnerable to the effects of a drug when the response is tested at a high drive level than when it is tested at a low drive level? An answer to this simple empirical question is likely to contribute to an understanding of the way in which drive (*D*) combines with other variables in determining response output, and might clarify the results of the studies (see Hulicka, 1960, for review) of the interaction of *D* with habit strength (*H*) and reward (*K*).

In the present investigation, the combinations of each of several levels of drive (thirst) with two drugs were studied. Variations in thirst drive were obtained by "prefeeding" *S*s different amounts of water before testing. The drugs employed were Chlorpromazine, a mild depressant ("tranquilizer"), and methylphenidate, an amphetamine-like stimulant. Though, like other drugs described as stimulants, methylphenidate, in medium doses of from 2 to 10 mg/kg, increases general activity (Bindra & Baran, 1959), we have repeatedly observed that it decreases the rate of a water-, food- or saccharin-rewarded lever pressing response; in this respect methylphenidate resembles amphetamine, which is also known to decrease such responding under certain conditions (e.g., Hearst, 1961; Miller, 1956; Owen, 1960). By using doses of Chlorpromazine and methylphenidate that would have roughly equal decremental effects, we hoped to study the interactions of drive with these drugs at comparable levels of response output.

METHOD

Subjects.—Twenty-eight naive male hooded rats, about 90 days old (weighing between 170 and 205 gm.) at the beginning of the experiment, were used as *S*s. They were housed in small metal cages, 2 to a cage. Food was available to them at all times except during the experimental sessions. The *S*s were placed on a 23-hr. schedule of water deprivation the day after their arrival at the laboratory, and they remained on this schedule for the duration of the experiment. They were allowed access to water in their home cages for 40 min. daily, at the conclusion of the experimental session.

Apparatus.—The apparatus consisted of four identical lever boxes, made of plywood and measuring 7 in. wide, 8¼ in. deep, and 11½ in. high. All the boxes were painted gray on the inside and were covered by a wire mesh roof. Each was equipped with a water-delivery mechanism and an automatic counter. A lever, measuring 4 × ¾ in., was mounted 3 in. above the floor on one of the smaller walls of each box. A tray for receiving water was located on the same wall, 2¼ in. from the center of the lever and 2 in. above the floor. Each lever press delivered .02–.04 cc of water into the tray. The minimum weight required to depress the lever was 44 gm.

Procedure.—After 1 day of water deprivation, each *S* was placed in a lever box for 10 min. The trays were filled with water; by the end of the period most *S*s had begun to drink the water. A 13-day training period was initiated on Day 2. On Days 2–9 each *S* was placed in a lever box for 15 min.; on Days 10–14, for 10 min. A given *S* was always placed in the same one of the four lever boxes. At the end of the training period the 4 *S*s with the lowest lever pressing rates were discarded, leaving 24 for the tests.

Variations in thirst drive were produced by allowing each *S* to drink a predetermined amount of water during a 10-min. period beginning about 40 min. before the daily test

TABLE 1

SUMMARY OF EXPERIMENTAL PROCEDURE

Stage	"Predrinking" in cc			Drug Condition	
	High Drive	Medium Drive	Low Drive	Drug Injection	Water Injection
1	0.0	2.5	5.0	1.5 mg/kg Chlorpromazine	0.3cc
2	2.5	7.5	10.0	1.5 mg/kg Chlorpromazine	0.3cc
3	0.5	4.0	8.0	4.0 mg/kg methylphenidate	0.75cc

session. Each *S* was placed with its allotted quantity of water in a box measuring 7 × 8 in. Except when the highest amount (10 cc) of water was given, all *S*s drank all their allotted water in the given 10-min. period. When *S* did not finish the allotted water, it was placed in its home cage and given an additional 10-min. period to finish the water; at the conclusion of this period there was never more than 1 cc of water left over.

Immediately after water consumption, that is, about 30 min. before testing, each *S* was weighed and injected intraperitoneally with an appropriate amount of distilled water (control), Chlorpromazine, or methylphenidate. The test sessions were conducted in three stages, each consisting of six sessions, one per day, with one practice session in between the stages. During each stage, each *S* was tested twice under each of the three drive levels. The two test sessions at each drive level were conducted on consecutive days; one of the tests followed a drug injection, while the other followed a control injection. The order of the three drive levels and the two types of injection (drug vs. control) was varied so that 2 *S*s were tested in each of the 12 possible permutations of drive level and type of injection. Thus, each of the 24 *S*s was tested under the six conditions (3 drive levels × 2 drug conditions) in each of the three stages of the experiment. The exact conditions of drive and drug employed in each stage are summarized in Table 1. It may be seen that, in Stages 1 and 2, *S*s were tested under the influence of 1.5 mg/kg Chlorpromazine at five different drive levels; in Stage 3, effects of methylphenidate at three drive levels were determined.

RESULTS

The mean numbers of responses made in the 10-min. test sessions under the different conditions of drive and drug are shown in Fig. 1 and 2. The higher scores obtained in Stages 2 and 3 can be accounted for by the fact that *S*s had not been trained to an asymptote before the initiation of the test sessions, so that their perform-

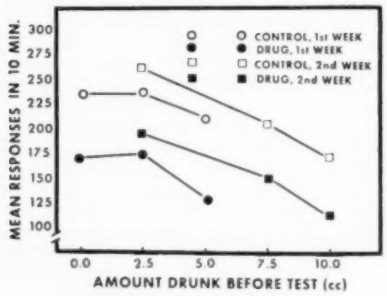

FIG. 1. Number of lever presses as a function of five drive levels (high to low) following control (water) and drug (Chlorpromazine, 1.5 mg/kg) injections.

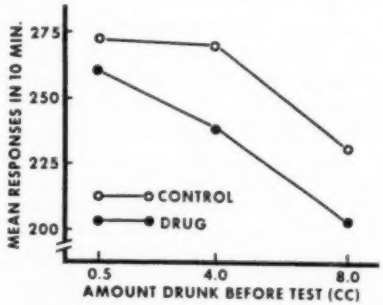

FIG. 2. Number of lever presses as a function of three drive levels (high to low) following control (water) and drug (methylphenidate, 4 mg/kg) injections.

TABLE 2

VALUES OF t FOR DIFFERENCES BETWEEN MEANS AT DIFFERENT DRIVE LEVELS

Stage	Control Scores			Drug Scores		
	D_L-D_M	D_M-D_H	D_L-D_H	D_L-D_M	D_M-D_H	D_L-D_H
1	2.72*	0.04	2.39*	3.35**	0.46	2.34*
2	3.93**	4.03**	9.29**	2.61*	2.50*	5.20**
3	3.54**	0.31	3.85**	2.71*	1.48	4.30**

* $P < .05$.
** $P < .01$.

ance was still improving during the test sessions. Three features of the curves in these figures should be noted. First, in the control, water-injection, conditions, the higher the drive level (i.e., the less the amount of "predrinking") the greater were *S*s' response rates. Two-tailed t tests for differences between correlated means indicated that most of the differences between the means at different drive levels were significant (see Table 2); the only nonsignificant differences were those between medium and high drive levels in Stages 1 and 3. Second, the level of performance in the drug conditions was also different at the different drive levels; the relation between drive level and response rate was the same as that exhibited under control conditions—the higher the drive level, the greater the rate of response. Again most of the differences were significant (see Table 2); as in the case of control scores, the differences between the scores at the medium and high drive levels in Stages 1 and 3 were not significant. Third, both drugs led to a decrease in the number of lever presses; the differences between the mean drug and mean control performances were significant at each drive level ($P < .05$), except under the high drive methylphenidate condition.

To determine whether the effects of the drugs on response rate varied with drive level, two measures of response decrement were calculated for each *S*: (*a*) the absolute difference between drug and control scores, and (*b*) the ratio of drug score to control score (D/C). The latter measure was used to minimize the effects of the absolute level of performance on response decrement. Using the absolute measure, the t test for correlated means indicated no significant differences between the means of absolute decrement in response scores at the different drive levels; this was true for both drugs. Using the ratio measure (Table 3) it was also found that the mean ratio of drug to control performance was independent of the drive level at the time of test. In the case of Chlorpromazine, the ratios at the lowest drive level were somewhat smaller than those at the two higher drive levels, but this apparent trend is not reliable.

DISCUSSION

Under the conditions of the present experiment, mild doses of Chlorpromazine and methylphenidate decreased the rate of lever pressing for water. The decremental effect of methylphenidate requires a word of comment. Even a cursory glance at the results of studies of the effects of stimulants on a lever

TABLE 3

MEANS AND *SD*s OF RATIOS OF DRUG TO CONTROL PERFORMANCE AT DIFFERENT DRIVE LEVELS

Stage	Drive Level					
	Low		Medium		High	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	.58	.31	.73	.28	.75	.22
2	.60	.36	.79	.30	.76	.28
3	.94	.28	.89	.29	.97	.19

pressing response (e.g., Hearst, 1961; Miller, 1956; Owen, 1960; Skinner & Heron, 1937; Teitelbaum & Derks, 1958) shows these effects to be quite variable; incremental effects are found together with decremental effects, sometimes in the same experiment. We believe that these discrepancies have arisen mainly from three types of differences in procedure: (*a*) Dose level: in general, the higher the dose the more likely it is that the stimulant will produce a decremental effect. (*b*) Schedule of reinforcement: continuous reinforcement schedules are more likely to yield decremental effects; schedules, such as partial and extinction, under which the animal makes many multiple responses (lever presses without entering the goal area) are likely to yield incremental effects. (*c*) Distance between the lever and the reward area: in general, the greater this distance, the more likely it is that the stimulant will decrease response rate. (It may be that the crucial factor here is the size of the lever box rather than the distance between the lever and the reward area). Using the specifications of the procedure employed in the present experiment, we have consistently obtained decremental effects with from 2 to 10 mg/kg of methylphenidate. As methylphenidate, like amphetamine, has an anorexic effect (Karczmar & Howard, 1959), it cannot be said whether the obtained decrease in lever pressing rate arises from anorexia or from interference caused by increased general activity; however, we are inclined to attribute greater importance to the latter factor (Bindra, 1961).

The decremental effects of both Chlorpromazine and methylphenidate, as measured by the absolute and proportionate decrement in response scores in the present study, were fairly constant over a wide range of drive levels. For Chlorpromazine, the effects were constant over the entire range of drive levels used (see Fig. 1 and Table 3). For methylphenidate, although the effects at different drive levels were not significantly different from each other, the absolute decrement in response rate was significantly different from zero at the two lowest drive levels, but not at the highest drive level (see Fig. 2 and Table 3). Our casual observations suggested one possible basis for the difference between Chlorpromazine-drive and methylphenidate-drive interactions. It seemed to us that while the decrement in the response rate under the influence of Chlorpromazine resulted from over-all slowing of *S*'s movements, the methylphenidate-induced decrement was associated with an increase in general activity, mainly repeated sniffing of different parts of the apparatus. Whether this difference is in any way related to the observed differences in interaction effects is a question that we are now investigating.

Payne (1958), in extending Hullian theory to account for response decrements resulting from drugs, concluded that "manipulations resulting in higher reaction potential will, under constant dosage, attenuate the degrading effect of a drug" (p. 342). According to Payne's analysis, one would expect the depressing effects of drugs to be less when the response is tested at high drive than when it is tested at a low drive level. The present results show this to be true only in the case of methylphenidate. Apparently, the simple relation suggested by Payne has limited applicability. The factors that determine the exact manner in which various drive and drug effects combine remain to be determined.

Summary

Is the effect of the injection of a given drug on a response dependent upon the drive level at which the response is tested? The effects of injections of Chlorpromazine (1.5 mg/kg) and methylphenidate (4 mg/kg) on a water rewarded lever pressing response were studied at several levels of thirst. Within three phases of the experiment all *S*s were tested, in a balanced order, under each of the six conditions (3 drive levels×2 drug conditions).

The following results were found: (*a*) The higher the drive level, the greater the response rate in both the control and the drug conditions. (*b*) Both Chlorpromazine and methylphenidate decreased response rate, though the former drug is a mild depressant

("tranquilizer") and the latter a stimulant. (*c*) There were no significant differences between the mean absolute, or relative, decrements in response score produced by either of the drugs at the different drive levels. (*d*) Methylphenidate failed to produce any marked decrement at the highest drive level; thus the response was less vulnerable to the effects of this drug at a high drive level than at low drive levels.

The decremental effects on the response of stimulants, such as methylphenidate and amphetamine are discussed. Some suggestions are presented regarding the basis of the difference in methylphenidate-drive and Chlorpromazine-drive combinations.

References

Bindra, D. Components of general activity and the analysis of behavior. *Psychol. Rev.*, 1961, **68**, 205–215.

Bindra, D., & Baran, D. Effects of methylphenidylacetate and Chlorpromazine on certain components of general activity. *J. exp. Anal. Behav.*, 1959, **2**, 343–350.

Hearst, E. Effects of d-amphetamine on behavior reinforced by food and water. *Psychol. Rep.*, 1961, **8**, 301–309.

Hulicka, I. M. Combination of drive and incentive. *Quart. J. exp. Psychol.*, 1960, **12**, 185–189.

Karczmar, A. G., & Howard, J. H. Anorexigenic action of methylphenidate (Ritalin) and Pipradrol (Meratran). *Proc. Soc. Exp. Biol. Med.*, 1959, **102**, 163–167.

Miller, N. E. Effects of drugs on motivation: The value of using a variety of measures. *Ann. N. Y. Acad. Sci.*, 1956, **65**, 318–333.

Owen, J. E., Jr. The influence of dl-, d-, and l-amphetamine and d-methamphetamine on a fixed-ratio schedule. *J. exp. Anal. Behav.*, 1960, **3**, 293–310.

Payne, R. B. An extension of Hullian theory to response decrements resulting from drugs. *J. exp. Psychol.*, 1958, **55**, 342–346.

Skinner, B. F., & Heron, W. T. Effects of caffeine and benzedrine upon conditioning and extinction. *Psychol. Rec.*, 1937, **1**, 340–346.

Teitelbaum, P., & Derks, P. The effect of amphetamine on forced drinking in the rat. *J. comp. physiol. Psychol.*, 1958, **51**, 801–810.

(Received April 3, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 510-513

ANAGRAM SOLUTION TIMES: A FUNCTION OF WORD TRANSITION PROBABILITIES

M. S. MAYZNER AND M. E. TRESSELT

University of Southern California *New York University*

Mayzner and Tresselt (1958, 1959) demonstrated anagram solution times to be a function of three variables: (*a*) anagram letter order, (*b*) word frequency, and (*c*) transition probability. The first variable, anagram letter order, refers to the rearrangement of the letters from their correct position in the word, while the second variable, word frequency, refers to the frequency of occurrence in English of the anagram solution as estimated from Thorndike-Lorge (1944) counts. The third variable, transition probability, refers to the probability or frequency with which given letters follow or precede other letters and digram frequency tables are available (Pratt, 1942; Underwood & Schulz, 1960) which permit the relative assignment of a numerical value to this frequency. For this variable it was found that anagrams whose summed transition probability or digram frequency totals are low yield significantly faster solution times than anagrams whose summed transition probability totals are high.

The present study examines the effect of word transition probability on anagram solution times, by selecting anagrams whose solutions are words yielding either relatively low summed digram frequency totals or relatively high summed digram frequency totals. It is predicted that anagrams whose word solutions have high transitions or digram totals will be solved significantly faster than anagrams whose word solutions have low transition totals, since it is assumed that *S*'s successive rearrangements of the letters of the anagram are more likely initially, at least, to correspond to sequential letter pairs with high transition probability or digram frequencies than low (e.g., the letters H and E are more likely to be arranged as HE than EH) and therefore more likely match words with high transition probability totals than words with low transition totals.

METHOD

Subjects.—The *S*s were 25 students from introductory psychology courses at Barnard College, Columbia University.

Stimulus material.—The stimulus material consisted of 18 five-letter anagrams typed in caps on white 3 × 5 in. cards. The 18 words from which the anagrams were constructed were selected from the Thorndike-Lorge (1944) word list on the basis of two criteria: (*a*) all words occur 100 times or over per million; (*b*) 9 words have low summed transition probability or digram frequency totals and 9 words have high summed transition totals. Summed transition totals for each word were obtained by adding the digram frequency values, as given by Underwood and Schulz (1960), for the four sequential letter pairs in each word. For example, with the word JUDGE, the four digram frequency values for the letter pairs JU, UD, DG, and GE are 30, 29, 18, and 201, respectively, yielding a summed digram frequency total for the word of 278, while with the word MONTH, the letter pairs MO, ON, NT, and TH yield values of 204, 786, 511, and 1795, respectively, for a summed total value of 3296.

All words and anagrams employed in the study may be found in Table 2. Digram frequency totals for the words range from 278 to 938 for the low transition probability (TP) words with a mean TP value of 725, and from 1647 to 3296 for the high-TP words with a mean TP value of 2190. Since it has been shown that the summed TP total of anagrams affects solution time (Mayzner & Tresselt,

1959), the anagram TP totals were made essentially equivalent for the low- and high-TP words, with a mean anagram TP value of 159 associated with the low-TP words and a mean anagram TP value of 161 associated with the high-TP words. The "similarity" of anagram to word letter order also was minimized in all 18 anagrams. Thus, anagram letter order, word frequency, and anagram TP totals were made essentially equivalent for all 18 words, allowing only the variable under investigation, namely, word transition probability to vary from low to high values.

Procedure.—Each of the 25 *S*s received the 18 anagrams in a different random order, with the constraint that the first, middle, and last 6 anagrams, each contain 3 anagrams whose solutions were high-TP words and 3 anagrams whose solutions were low-TP words. All *S*s, who were tested individually, were told that a five-letter anagram is simply five scrambled letters which, when properly rearranged will make a single five-letter word. After being shown an example, they were informed that they would have 18 anagrams to solve and if they could not find a solution within 2 min. the correct solution would be given to them before going on to the next anagram. Solutions were given by *S* verbally (i.e., *S* had no paper or pencil) and solution times were recorded with a stop watch.

Results and Discussion

Available for analysis were 18 time scores for each *S*, with each score having a possible range from 1–120 sec. Since each *S* received nine high-TP words and nine low-TP words, the medians of each of these two sets of nine scores for each *S* were used as the criterion measures for statistical evaluation. Medians were employed rather than means because of skewed distributions and the artificial ceiling imposed by the 120-sec. time limit per anagram. The means, *SD*s, medians, and ranges of these criterion measures are presented in Table 1, and the results indicate that anagrams whose word solutions have high-TP totals are solved significantly faster than anagrams whose word solutions have low-TP totals, with the correlated t yielding a value of 2.70 ($df = 24$; $P < .02$). If the sign test is applied to the 25 pairs of scores, 17 pairs are in the predicted direction, 5 pairs opposed, and there are two ties ($P < .01$).

An analysis also was made for each of the 18 individual words. These

TABLE 1

Anagram Solution Times (in Sec.) for High and Low Transition Probability Conditions (Based on Medians)

Transition Probability	Mean	*SD*	Median	Range
High	14.40	10.95	12.00	3–47
Low	28.00	30.03	16.00	3–120

TABLE 2

Anagram Solution Times (in Sec.) for Individual Words with High and Low Transition Probability Totals

High-TP Words	Anagrams	Median	Range	Low-TP Words	Anagrams	Median	Range
WATER	AEWTR	3	1–56	JUDGE	EGUJD	3	1–45
HOUSE	EUOHS	6	1–120	VOICE	EOCVI	4	1–66
DRINK	NRDKI	7	1–83	GUIDE	IUEGD	7	2–120
PAINT	IPTNA	13	2–120	FRUIT	IUFTR	15	1–120
MONTH	OHTNM	13	3–120	MUSIC	IUMCS	27	2–120
HUMAN	MHNUA	15	2–120	VALUE	AEUVL	46	2–120
POUND	UODNP	17	2–120	GUARD	AUGDR	49	1–120
YOUTH	OYHTU	18	1–120	UNCLE	EUCNL	72	1–120
POWER	OEWRP	22	2–94	SUGAR	SRGUA	83	2–120

results are presented in Table 2, which gives the medians and ranges of solution times based on the performance of the 25 *S*s. While five of the low-TP words have longer median solution times (i.e., 27, 46, 49, 72, and 83 sec.) than any of the high-TP words, whose median solution times range from 3 to 22 sec., four low-TP words fall within this range (i.e., 3, 4, 7, and 15 sec.).

Since a previous study by Jenkins and Mayzner (1961) found a significant reduction in anagram solution times over the course of 10 problems, a similar analysis was performed on the present data. No systematic change in solution times was found over the 18 problems. The discrepancy between the two studies is probably a function of the use of a restricted set of anagram letter orders in the previous study, while in the present study each anagram employed a different letter order.

In previous studies Mayzner and Tresselt (1958, 1959) suggested that an anagram constitutes a problem whose stimulus properties tend to evoke a variety of implicit responses on the part of *S*, and letter order, word frequency, and anagram transition probability were related to the way in which these implicit responses were produced. By conceptualizing anagram problem solving in this way, one is, in fact, employing an S-R mediational model of the type being extensively employed by Kendler and his associates (Kendler & D'Amato, 1955; Kendler & Karasik, 1958; Kendler & Mayzner, 1956) in treating concept formation processes. The present study extends this mediational analysis to include the effect of word transition probability on anagram solution times and the results indicate that the model is successful in making predictions regarding the effect of this variable.

While the four variables so far studied appear to be critical factors governing the type of implicit responses produced, they apparently far from exhaust the variance associated with anagram solution times. This is most evident in the wide range of solution times, from 3 to 83 sec., found with the low-TP words. The words GUIDE and GUARD provide an example of this unaccounted for variance, in that the solution time of GUARD (49 sec.) is seven times that of GUIDE (7 sec.), even though four variables have been held constant and three of the five letters are the same in both words. Such findings suggest that further research is required to uncover the variables involved in the production of implicit responses to anagram problems.

Summary

A study was made of the effect of word transition probability on anagram solution times. Word transition probability was varied by selecting words with either relatively high or low summed digram frequency totals, while holding constant anagram letter order, word frequency, and anagram transition probability variables. The results, predicted on the basis of an S-R mediational model, showed that anagrams whose word solutions have high transition probability totals are solved significantly faster than anagrams whose word solutions have low transition totals.

References

JENKINS, H. M., & MAYZNER, M. S. The time-pattern of switching among alternatives in problem solving. *Psychol. Rep.*, 1961, **8**, 419–430.

KENDLER, H. H., & D'AMATO, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.

KENDLER, H. H., & KARASIK, A. D. Concept formation as a function of competition between response produced cues. *J. exp. Psychol.*, 1958, **55**, 278–283.

KENDLER, H. H., & MAYZNER, M. S. Reversal and nonreversal shifts in card-sorting tests with two or four sorting categories. *J. exp. Psychol.*, 1956, **51**, 244–248.

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of letter order and word frequency. *J. exp. Psychol.*, 1958, **56**, 376–379.

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of transition probabilities. *J. Psychol.*, 1959, **47**, 117–125.

PRATT, F. *Secret and urgent.* New York: Blue Ribbon, 1942.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Columbia University, Bureau of Publications, 1944.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

(Received April 7, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 514–518

TRANSITION PROBABILITY EFFECTS IN ANAGRAM PROBLEM SOLVING[1]

HARRY BEILIN AND RHEBA HORN

Brooklyn College

It has been proposed that if certain gestalt assumptions about perceptual and cognitive organization are true, meaningful letter groups such as words should resist reorganization more than meaningless letter aggregates when these are the materials of anagram problem solving experiments (Hollingworth 1935, 1938). Hollingworth and others (Nissenson & Sargent, 1941; Sargent, 1940), however, have found words to be no different from nonsense letter arrangements in resisting alteration. Up to the time of Johnson's (1955) review only Devnich's (1937) findings suggested that a difference exists. This was sufficient for Johnson to consider the question still open.

In the main, the technique used in these experiments required that *S* discover the word (e.g., KANGAROO) that could be made from either related or unrelated word groups (e.g., AGO, KORAN) in contrast with meaningless letter arrangements (e.g., OAG, KRNOA). More recently, Mayzner and Tresselt (1958, 1959) have tested the aforementioned thesis employing meaningless letter aggregates as anagrams. They establish that anagram solution time is a function of anagram letter order, solution word frequency (Mayzner & Tresselt, 1958), and transition probability (Mayzner & Tresselt, 1959). By contrasting the solution times of anagrams of high transition probability total (from summed digram frequencies of sequential letter pairs) with those of low transition probability total a positive relationship is established between the two variables. In other words, the higher the transition probability total of a letter aggregate the more difficult it is to alter the letter arrangement and the longer the solution time. Although Mayzner and Tresselt (1959) do not directly infer from these results that the difficulty of altering letter arrangements in words, in contrast to nonsense letter arrangements, is a function of transition probability they do conclude that well patterned letter groups (which they illustrate with words) are more difficult to reorganize than loosely patterned letter groups (which they illustrate with nonsense arrangements) if patterning is defined in terms of transition probabilities.

In the foregoing experiments two variables are considered and to some extent confounded. Words and nonsense letter arrangements used as anagrams differ both in meaning and organization (or patterning). These variables require independent treatment since it is not at all difficult to have nonsense letter arrangements which are more highly patterned than words (if patterning is taken as the equivalent of transition probability total, as Mayzner and Tresselt propose). In the Mayzner and Tresselt studies only nonsense letter arrangements of greater and lesser patterning (transition probability totals) are employed. It is dubious therefore whether the transition probability hypothesis can be applied as an explanation for solution time differences between word and nonsense anagrams, without a test in which words and nonsense arrangements are employed.

The present experiment offers a test of the hypothesis that solution time differences between highly and loosely patterned anagram letter arrangements are a function of transition probability. This is achieved by utilizing both five-letter words and nonsense five-letter aggregates as anagrams, keeping transition probability, letter order, and solution-word frequency constant.

[1] Presented at the 1961 meetings of the Eastern Psychological Association in Philadelphia. We express our appreciation to Max Weiner for his assistance in making *S*s available to us. Part of this study was an honors project by Rheba Horn.

METHOD

Subjects.—Sixty students, primarily in their junior year, at Brooklyn College served as *Ss*; 30 were randomly assigned to each of two groups which differed in the presentation order of the materials.

Materials.—The stimulus materials consisted of 10 five-letter nonsense aggregates (e.g., OBAVE) typed in capital letters on plain 3 × 5 in. index cards, and another set of 10 five-letter words (e.g., FROTH) typed in identical fashion on index cards. Both words and nonsense letter aggregates are used as anagrams in the experiments.

Experimental design.—One group was first given the 10 nonsense letter anagrams and then the 10 word anagrams; the other group the reverse. The presentation order of individual anagrams was rotated so that every anagram arrangement occurred in each serial position and in a different combination with its counterpart in the other anagram group (i.e., 1,1; 2,3; 3,5; . . . 10,10).

The solution words for both nonsense and word anagrams were selected from the Thorndike-Lorge (1944) word list and equated for frequency. The mean frequencies of nonsense and word anagram solution words are 74.4 and 74.3, respectively.

The traditional anagram problem was modified so that only *two* letters of the nonsense or word anagram were to be interchanged to produce the solution word (keeping the remaining letters in order). The "letter switching order" of the nonsense and word anagram sets was matched to control letter order alteration (i.e., there is one 1–3 switch in each group, one 1–4, one 1–5, etc). This is necessary since some letter orders are more difficult than others (Hunter, 1959); limiting the altering task to two letters controls the unscrambling necessary to solve the anagram.

The anagram sets were also equated for transition probability. The nonsense anagrams were arranged to yield mean transition probability totals equal to the mean totals for the word anagrams. Transition probability totals are derived from summed digram frequencies of sequential letter pairs (Pratt, 1942). For example, the total for the anagram OBAVE is the total of OB, BA, AV, and VE = 4 + 10 + 10 + 29 = 53. The mean transition probability total for nonsense anagrams is 153.8; for word anagrams, 151.7 (the difference is not significant).

The use of transition probability totals from summed digram frequencies is based upon *S*'s supposed knowledge (of probabilities of occurrence) of pairs of letters in his native language. Since *S* may be equally affected by his knowledge of three-letter, four-letter, and five-letter groupings and single letter expectancies in his attempts at solution it suggests that the total transition probability of an anagram should be a composite of digram, trigram, quadragram, and pentagram frequencies, as well as initial and terminal letter frequencies. Pentagram and quadragram frequency tables were not available to us, although trigram frequency tables were (Pratt, 1942). We have computed the mean total of summed trigram frequencies for word and nonsense anagrams. The mean for word anagrams is 204.2 (SD = 392.2); for nonsense anagrams, 164.8 (SD = 355.8). The difference between these means is not significant. In addition the difference of mean initial letter frequencies is not significant (word mean = 63.0, SD = 63.1; nonsense mean = 67.4, SD = 58.5), nor the difference between terminal frequency means (words mean = 122.6, SD = 72.5; nonsense mean = 72.7, SD = 81.7).

Procedure.—The *Ss* were told they were to make meaningful words from the nonsense and word anagrams by interchanging two of the letters and keeping the remaining three in the same location. Only one correct solution is possible for each anagram. One example (i.e., CAUSE-SAUCE, ERTEN-ENTER) and one sample problem (i.e., SPOON-SNOOP, FALSH-FLASH) was given before each anagram list. The *Ss* were also told there would be 10 anagrams in each list with a maximum 2-min. solution time per anagram, after which the solution word would be given before continuing to the next anagram. Solutions were given by *S* verbally with no aids permitted. Solution times were recorded with a stop watch. Testing was individual and the experiment proper was followed by asking *Ss* how they obtained the solutions.

RESULTS AND DISCUSSION

Medians of *S*'s solution times for each anagram set and median group (N = 30) solution times per anagram were computed. Table 1 gives the major results expressed as the means and SDs of individual medians. Within this design it is possible to evaluate the effects upon solution time of the word-nonsense letter arrangement differences, as well as possible practice

TABLE 1

SOLUTION TIMES (SEC.) FOR NONSENSE ANAGRAM AND WORD ANAGRAM TASKS

Presentation Order	Word Anagrams			Nonsense Anagrams			Total		
	N	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*
First	30	18.23	21.45	30	8.37	11.19	60	13.30	17.68
Second	30	15.17	12.01	30	10.46	12.75	60	12.81	12.41
Total	60	16.70	17.33	60	9.42	11.83			

Note.—Means and *SD*s are based upon medians of *S*'s solution times for 10 nonsense anagrams and 10 word anagrams.

effects. The word-nonsense arrangement effect was tested with the Friedman two-way analysis of variance by ranks (Siegel, 1956). Word anagram solution time was significantly slower than nonsense anagram solution time ($P < .001$). (First and second order presentations were combined, on the basis of the analysis which follows.)

The effects of practice were assessed separately for the word and nonsense letter arrangement groups. A Mann-Whitney U test was applied to the first and second order solution times within each task. The values indicated are not significantly different ($P > .10$).

To assess further the relationship between transition probability and solution time (since transition probability total is already controlled) Kendall τ's were computed between nonsense anagram solution times and their associated transition probability totals. The same was done for the word anagram solution times. The τ's for nonsense anagrams, first and second order presentations, were $-.06$ and $-.16$, respectively. For words, these τ's were .11 and .08. These correlations are not significant. On the other hand, τ between solution times for the first and second presentation of word anagrams was .62, and the comparable τ for nonsense anagrams was .63 ($P < .05$).

The anagram problem solver has two kinds of data to take into account: the anagram and the solution words he generates. If he requires knowledge of the transition probabilities of the letter pairs in his language to treat adequately the letter pairs of the anagrams (Hunter, 1959) the same should apply to the solution words. This is evident from those *S*s who put correct combinations together but do not recognize them as words. (We might be tempted to say that the process of solving anagrams is one of testing varieties of possible letter combinations and then recognizing the combinations as words, were it not that many *S*s produce solution words and *then* check the correctness of the letter combinations.) We tested to be certain the transition probability totals for *solution* words were not different, even though word frequencies were equated. The difference between averaged transition probability totals (from summed digram frequencies) for solution words is not significant (word mean = 145.1, $SD = 63.9$; nonsense mean = 151.4, $SD = 75.3$).

The *S*s divide into those who solve nonsense anagrams faster and a smaller group who solve word ana-

grams faster. To test whether these represent stable *S* capacities, we first hypothesized that the *S* trait represented here is the ability to deal with embedded figures (viz., letters embedded in words). Such an *S* capacity might be related to an ability to deal with geometric figures embedded in complex designs as represented by the Gottschaldt figures.[2] No such relationship, however, was found to exist.

As Mayzner and Tresselt (1958) and Hunter (1959) show, by different methods, the production of solutions per unit time comes very quickly at first and then tapers off. By applying the method used by Hunter, which indicates the average rate of correct solution production, to a 10-sec. interval, we find the same solution production pattern except that the production rate for word anagrams is slower than for nonsense anagrams. Examination of *S*s' retrospective reports suggests why this may be so. Some *S*s comment that word anagrams often bind them whereas nonsense anagrams do not. These *S*s perseverate on the word anagrams to such an extent that problem solution is interfered with. Whether perseveration is due to attention to the meaning associated with the word or to its sound is not clear.

Word anagrams are more difficult to alter than nonsense anagrams. This supports the findings of Mayzner and Tresselt (1958) and Devnich (1937). Since transition probability was controlled the Mayzner and Tresselt explanation of the difference may be rejected. The low correlation between solution time of individual anagrams and their transition probability totals further supports such rejection. Although solution time differences due to patterning may be accounted for by transition probability it does not follow that it will similarly account for differences in meaning.

Although it has not been possible to compute composite transition probability values for the anagrams, we may conclude that anagram solution time differences between words and nonsense arrangements are not due to averaged digram, trigram, or initial and terminal letter frequencies.

Two possibilities remain to account for the anagram solving differences. The first is the perseverative effect generated by word anagrams and the second, organization or patterning effects independent of transition probabilities of successive letter arrangements.

[2] We therefore administered the Thurstone Concealed Figures Test (Thurstone, 1944, 1951) on a group basis to 40 of the 60 *S*s still available to us. The mean correct solutions were 81.9 for nonsense anagrams ($SD = 15.5$) and 84.2 for words ($SD = 17.2$) ($P > .05$). We are indebted to Aaron Carton for suggesting this possibility to us.

Summary

An experiment tested the effect of transition probability of successive letter sequences upon the solution time of word and nonsense anagrams. Transition probability totals (from summed digram frequencies of sequential letter pairs), solution word frequencies, and letter switching orders were equated for word and nonsense anagrams.

Solution time is significantly longer for word anagrams than nonsense anagrams. This difference is not accounted for by transition probability totals from summed trigram frequencies, initial and terminal letter frequencies, or transition probability totals of solution words.

Word perseveration effect is offered as a possible explanation of the solution time difference as well as organization effects not accounted for by transition probabilities of successive letter arrangements.

References

Devnich, G. E. Words as "Gestalten." *J. exp. Psychol.*, 1937, **20**, 297–300.

Hollingworth, H. L. The conditions of verbal configuration. *J. exp. Psychol.*, 1935, **18**, 299–306.

HOLLINGWORTH, H. L. Verbal Gestalt experiments with children. *J. exp. Psychol.*, 1938, **23,** 90–95.

HUNTER, I. M. L. The solving of five-letter anagram problems. *Brit. J. Psychol.*, 1959, **50,** 193–206.

JOHNSON, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of letter order and word frequency. *J. exp. Psychol.*, 1958, **56,** 376–379.

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of transition probabilities. *J. Psychol.*, 1959, **47,** 117–125.

NISSENSON, M., & SARGENT, S. S. Words as configurations. *J. exp. Psychol.*, 1941, **28,** 85–89.

PRATT, F. *Secret and urgent.* New York: Blue Ribbon, 1942.

SARGENT, S. S. Thinking processes at various levels of difficulty. *Arch. Psychol., N. Y.*, 1940, No. 249.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Columbia University, Bureau of Publications, 1944.

THURSTONE, L. L. *A factorial study of perception.* Chicago: Univer. Chicago Press, 1944.

THURSTONE, L. L. An analysis of mechanical aptitude. *Psychometr. Lab. Bull.*, No. 62. Chicago: Univer. Chicago Press, 1951.

(Received April 19, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 519–520

SUPPLEMENTARY REPORT: A NONVERBAL MEASURE OF EXTINCTION IN SKILL AND CHANCE SITUATIONS[1]

KENNETH B. HOLDEN[2] AND JULIAN B. ROTTER

Ohio State University

James and Rotter (1958) have shown that partial reinforcement does not lead to more resistance to extinction than 100% reinforcement if *S* feels that reinforcements are produced as a consequence of his own skill. The James and Rotter study used verbal expectancies as a measure of learning and extinction. The present study is an attempt to expand the generality of the Internal-External (I-E) variable (Rotter, Seeman, & Liverant, in press) to include nonverbal behavior and therefore be more analogous to the classical extinction studies. This study is essentially a replication of the 50% reinforcement groups of the James-Rotter (1958) study with a betting technique as a behavioral measure of extinction rather than stated verbal expectancies. In addition, a 50% reinforced group receiving neither skill nor chance instructions, but rather instructions neutral on the I-E variable was included.

Method.—The method, except for the betting technique as a measure of extinction, was identical to that used by James and Rotter (1958). The apparatus, stimulus material, and other details of the experiment with the exception of the statement of verbal expectancies were duplicated, with the verbal expectancies being replaced by a money reward procedure. James and Rotter found that the majority of the *Ss* responded differentially to skill and chance instructions in a simple card-guessing game, although there were wide individual differences.

The *Ss* were seated about 5 ft. in front of the screen containing the shutter of the tachistoscope. Beside the screen was a pile of approximately 400 nickels with several $2.00 rolls of nickels alongside. The *Ss* were given $2.00 from this money at the start of the experimental procedure. A white 3 × 5 in. card was taped to the table. Before any cards had been shown, *E* placed a nickel from his pile on the card and *S* was requested to place one from his $2.00 on the card. After a set of seven cards, if *S* had been successful (i.e., had been allowed to correctly anticipate at least five out of seven cards), *E* added a nickel to the card and *S* removed one. If *S* had not been allowed to be successful *E* removed a nickel and *S* had to add one from his original $2.00 if he wished to continue playing. This procedure continued until *S* extinguished or ran out of money.

The James and Rotter (1958) instructions for the skill and chance groups were applied in this study, except for the difference in response measure used. The "ambiguous" group received all the instructions that were *common* to the skill and chance groups except for the statement in the first paragraph of the instructions that "There is some evidence that some people are considerably skilled at this."

The *Ss* were 38 men and 38 women, aged 27 yr. or less, from elementary psychology classes during the summer quarter. They were assigned to three groups in a random manner, with the restriction that there would be an equal number of men and women in each group.

Results.—The means and *SD*s for trials to extinction are given in Table 1. The measure was the number of trials (sets of seven card choices) taken by *Ss* to extinguish after an initial set of 10 learning trials. A maximum of 40 extinction trials was possible, with 2 of the 76 *Ss* failing to extinguish. A nonparametric analysis was used since the distributions did not meet the requirements for *t* tests.

The *SD*s are rather large, especially in the chance and ambiguous conditions, indicating the possibility that the instructions were not as effective for some *Ss* as they were for others and that some *Ss* operated according to their own hypotheses rather than to the instructions. An *F* test of the ratio of the variances between skill and chance conditions was 1.69, short of the 1.98 necessary for significance at the .05 level. The Mann-Whitney *U* for the difference between the skill and chance conditions gave a *z* of 2.78 (two-tailed $P = .005$). Additionally, the *z* for the difference between the chance and ambiguous conditions was 2.36 ($P = .018$). The *z* of .19 between the skill and ambiguous conditions was, of course, not significant.

It can be seen that, while men tend in the

[1] This research was aided in part by the use of funds and data collected for the United States Air Force Contract No. 49(638)-741 monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

[2] Now at San Fernando Valley State College, Northridge, California.

TABLE 1

MEANS AND *SD*s OF NUMBERS OF TRIALS TO EXTINCTION

Conditions	Men			Women			All *S*s		
	N	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*
50% Skill	13	13.6	5.1	13	10.7	7.1	26	12.4	6.4
50% Chance	13	18.7	10.7	13	23.2	10.0	26	20.9	10.6
50% Ambiguous	12	17.2	11.1	12	11.2	7.9	24	14.1	10.1

same direction, much of the experimental effect appears to stem from women in the population sampled. The Mann-Whitney *U*s for women were 26 ($P = .002$) and 20 ($P = .002$), respectively, for the differences between the skill and chance and the chance and ambiguous conditions. The corresponding *U*s of 70 and 51 for the men were not significant at the .05 level. It will be noted that in the ambiguous condition the men responded as they did under the chance condition and the women responded as they did under the skill condition.

This study extends the generality of the James and Rotter (1958) findings by showing that the same effects of skill and chance instructions are obtained when a nonverbal response is used to index experimental extinction. There appears to be a sex difference in the effects of the instructions, with the ambiguous instructions giving the same effects as chance instructions with men but giving the same effects as skill instructions with women.

REFERENCES

JAMES, W. H., & ROTTER, J. B. Partial and 100% reinforcement under chance and skill conditions. *J. exp. Psychol.*, 1958, **55**, 397–403.

ROTTER, J. B., SEEMAN, M., & LIVERANT, S. Internal versus external control of reinforcements: A major variable in behavior theory. In D. Williams (Ed.), *Decisions, values, and groups*. New York: Pergamon, in press.

(Received December 8, 1960)

ERRATUM

In the article "The Effect of Stimulus Similarity on the Acquisition and Extinction of a Conditioned Response" by D. P. Hunt (*J. exp. Psychol.*, 1962, **63**, 278–283) the line in the fourth paragraph, third sentence (p. 278) should read as follows: "By assuming equal terminal levels of I_n. . . ."